

# THIS WEEK

## EDITORIALS

**FLOODS** The lessons New York needs to learn from Superstorm Sandy **p.148**

**WORLD VIEW** To misreport science should be viewed as misconduct **p.149**

**DISEASE** Bats not rats are most likely to pass on viral illness **p.150**



## A deal on the horizon

*Leaders have finally thrashed out the European Union budget for the next seven years. But how much money will go to research is yet to be confirmed.*

After almost 26 hours of intense debate last week, European leaders have finally agreed on the total European Union (EU) budget for the period 2014–20. Scientists can breathe a sigh of relief — but concerns cannot be dismissed just yet.

The deal allocates €125.6 billion (US\$168 billion) for initiatives to increase Europe's competitiveness and strengthen employment. That includes the budget for the Horizon 2020 research programme, which will fund basic research through the European Research Council (ERC) and applied science through other projects.

The total funding for competitiveness over the next seven years has increased by more than 37% compared with the EU budget for 2007–13. European scientists and research lobbyists have cautiously welcomed the deal, which hints at a reasonable settlement for research given Europe's current tight economic climate. As described on page 159, the deal currently sets aside around €69 billion for Horizon 2020. But that could change in coming weeks, as ministers thrash out the fine details of the agreement and the European Parliament also has its say.

The deal agreed on 8 February follows intense lobbying by scientists across the continent to protect the research budget after EU leaders failed to see eye to eye at budget talks in December. Hard-line governments including those of Britain and Germany were looking for a substantial slash to EU spending plans. Lobby groups including Euroscience, which is based in Strasbourg, France, and scientific leaders including Helga Nowotny, president of the ERC, urged decision-makers to safeguard the €80-billion research budget suggested by the European Commission.

Analysts have already begun to crunch the numbers to work out what the competitiveness budget could mean for research. According to estimates by Wolfgang Eppenschwandtner, executive coordinator of Initiative for Science in Europe (ISE), an independent science-advocacy group based in Heidelberg, Germany, the lobbyists will not get everything they wanted. The question is, on what will they lose out?

In the agreement, decision-makers said that a priority of EU spending should be to strengthen research and innovation. Horizon 2020 and the ERASMUS for All programme — which includes funding for graduate students to study abroad — have been promised more money in their yearly budgets than was provided for research and the ERASMUS programme in 2013.

Given this rhetoric, says the ISE, a research budget of €69 billion represents the worst-case scenario. In the best case, Horizon 2020 could actually be awarded between €75 billion and €78 billion.

Standing in the way of the best-case scenario are the financial commitments that EU leaders have already made from the 2014–20 competitiveness budget. They have set aside €29.3 billion for building transport, energy, broadband and digital-services infrastructure as part of the Connecting Europe Facility. They have also allocated €6.3 billion for the Galileo global-navigation satellite system; €2.7 billion for ITER, the experimental fusion reactor being built in Cadarache, France; and €3.8 billion for the Global Monitoring for Environment and Security

Earth-observation satellite. Taking into account all the other initiatives that the competitiveness budget must also cover, not all that much money is actually left in the budget for research.

One possible way to free up more funds would be for the EU to pay for its agriculture research, proposed at about €4 billion, from the agriculture budget instead of the competitiveness budget. Still, even

***“The lobbyists will not get everything they wanted. The question is, on what will they lose out?”***

lobbyists pushing for this idea say that it is not clear whether it is legal or possible given administrative constraints.

Ultimately, the final outcome for research depends on ministers' interpretation of the edict that Horizon 2020 will get a real increase over funding levels in 2013.

Keeping in mind that before the budget talks, rumours were circulating of even deeper cuts to research, scientists and lobbyists have done well to secure research funding of at least €69 billion. But the fight is not yet over. ■

## Preventive therapy

*Stem-cell trials must be made easier, so that treatments can be based on real data.*

Last November, a Nevada court convicted two men of fraud for selling ineffective stem-cell treatments to people chronically ill with, among other disorders, multiple sclerosis or cerebral palsy.

According to the US Food and Drug Administration (FDA), one of the men, Alfred Sapse, targeted extremely ill patients with a method that he claimed to be proprietary — implanting portions of placental tissue into the abdomen. Sapse, the agency says, knew that he needed FDA approval for such a procedure. He didn't have it. He claimed to be a doctor but didn't have a licence. The other defendant, the physician who performed the procedures at Sapse's bidding — on some 34 people in Las Vegas — knew “that it would not benefit the patients”. The pair “conducted no meaningful follow-up with the patients who underwent the implant procedures”. They did “not use any of the money for laboratory research, animal studies or human clinical studies relating to the short- and long-term effects of the implant procedures”. (Sapse made US\$1 million from the treatments; he spent \$700,000 of that on gambling and personal expenditure.) At least two patients suffered infections, and it is not clear what damage others might have incurred. In November 2006, the FDA issued a warning letter, telling the pair to stop. But they continued.

The incident shows the cavalier attitude with which many fraudsters

approach the promising yet immature field of stem cells. It also shows the importance of the FDA's regulatory role. Yet, as described on page 166, that role has been questioned by a Texas stem-cell-therapy company. The firm, Celltex Therapeutics in Houston, has demonized the agency to its patients, some of whom seethe at what they see as government intrusion.

The situation with Celltex is different from the Sapse case — the Texas company believes in its treatments, its doctors are real and its stem-cell manufacturing was registered with the FDA, for example. But the quality of the stem cells that Celltex used was not ensured, and follow-up on patients seems not to have been rigorous.

Celltex frames its dispute with the FDA as a conflict between a brave company that wants to offer cutting-edge medicine to desperate patients and a tyrannical bureaucratic ogre that is holding it back. But some of the facts don't fit that simple narrative.

Patient care is about more than intervention. It demands top-quality processing facilities, systematic administration of therapies and meticulous follow-up, so that informed decisions can be made for current and future patients. If Celltex wanted to be the standard-bearer for a new form of cell-based therapy, if it wanted to have a showdown with a federal agency by demonstrating that regulations stand in the way of scientific progress and patient health, it should have produced the best evidence of safety and efficacy that it could muster. That would have stimulated an interesting and constructive debate and created real pressure for change at the FDA.

Certainly, there is room for the FDA to improve the regulation of stem cells. The large clinical trials that the rules currently demand are so expensive that many researchers and biotechnology companies cannot afford to conduct them. To ease that problem, the agency could explore expanding its 'compassionate use' clause, which allows

individual patients to pay for drugs that are being used in FDA-approved trials. Alternative funding mechanisms, perhaps involving national insurance programmes, could be used to help offset the costs to patients and to those who perform the trials.

To enact such a change, the FDA would probably need more money to ensure that companies are serious about developing medicines and not simply seeking a loophole to increase profit. Increased FDA funding could also enable the agency to waive or reduce fees paid by the (usually small) biotechnology companies that develop the treatments. Organizations including the Alliance for Regenerative Medicine in Washington DC have been working with the FDA to create a regulatory environment more conducive to the development of stem cells.

More broadly, there has been a boom in people going abroad to receive stem-cell treatments. The World Health Organization (WHO) has taken a stand by issuing guidelines on how to regulate cell transplants. National authorities could actively engage with the WHO to ensure that those guidelines are effective.

All involved want to speed up the introduction of stem cells into the clinic. Patients are ready to take risks and clinical researchers are ready to do studies. Funding bodies should step up with money to help.

The matter is urgent. After the FDA turned up the heat, Sapse went to Mexico for three years before his arrest in 2010. Celltex is moving there now. The longer it takes to develop a workable and affordable system in nations such as the United States, the more patients will travel for treatment to countries where there are even more unknowns. ■

## Damage control

*Planning for extreme events must incorporate not just infrastructure but societal preparedness.*

When officials in New York City began to piece together how Superstorm Sandy had managed to flood the subway last October, they found that the storm had driven a bundle of lumber from a construction site right through a plywood barrier built around one of the entrances to the South Ferry subway station. It was a seemingly random act of violence, but in reality, the barriers probably never stood a chance. With a standing-water height of up to 1.5 metres at Battery Park on Manhattan's southernmost tip, the rising tide skirted a second plywood blockade and poured over a waist-high concrete wall at another entrance.

Preparing for hurricanes is hard. But the fact that core infrastructure in a global metropolis such as New York was protected by plywood should trigger alarms. South Ferry is a reminder of just how ill-prepared New York was for a storm of this magnitude — and it underscores the scale of the challenge ahead.

It wasn't supposed to be this way. New York City has engaged scientists while working to reduce emissions and prepare for a warmer world. In 2008, Mayor Michael Bloomberg created the New York City Panel on Climate Change, and in August the city council gave the panel a permanent place in its long-term planning process. *PlaNYC*, a planning document that offers a vision of what the city will look like in 2030, includes a comprehensive chapter on climate change. But none of this prepared the city for Sandy. Nor could it have — the surge that Sandy brought ashore was off the charts.

Legions of scientists are now assessing what happened and projecting future risks. The latest, and perhaps best, estimate, based on models by

researchers at Princeton University in New Jersey and the Massachusetts Institute of Technology in Cambridge, is that the storm surge at Battery Park was a 1-in-500-year event. But the size of a surge is not the only measure of a dangerous storm, nor is Battery Park the only location that matters. Scientists also know that the baseline is changing with the climate. All of which leaves the city, its residents and businesses in the unenviable position of rebuilding in the face of an uncertain future.

As this process unfolds, several lessons can be learned from Sandy (see page 162). In many places, premises erected under newer building codes survived the storm with only limited damage at ground level. A new generation of waterfront parks and developments also weathered the storm quite well, showing that there are ways to manage the risks of occasional flooding. But given the predicted sea-level rise and the likelihood of more powerful storms in the future, a more comprehensive strategy is clearly needed.

Some positive signs have emerged. The Federal Emergency Management Agency is updating the city's flood maps, and the city has announced steps to strengthen its building codes. As directed by Congress last year, the agency will also be incorporating long-term climate projections, including for sea-level rise, into its rate structure for the federal flood insurance programme. Until now, the programme has served as a government subsidy for risky coastal development — so risky that private insurance companies refused to enter the market.

One of the big questions facing the region is whether to spend billions of dollars on a storm-surge barrier. Scientists and engineers should clearly include a barrier in their analysis, but a surge is just one of many threats posed by many kinds of storm. Moreover, how fast New York bounces back will depend not only on damage to infrastructure but also on the strength of social networks and the general health of the communities affected. Farther afield, as sea levels rise, coastal cities will have little choice but to learn to live with more water than they are used to today. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunq](http://go.nature.com/xhunq)





## Redefine misconduct as distorted reporting

To make misconduct more difficult, the scientific community should ensure that it is impossible to lie by omission, argues **Daniele Fanelli**.

**A**gainst an epidemic of false, biased and falsified findings, the scientific community's defences are weak. Only the most egregious cases of misconduct are discovered and punished. Subtler forms slip through the net, and there is no protection from publication bias.

Delegates from around the world will discuss solutions to these problems at the 3rd World Conference on Research Integrity (wcri2013.org) in Montreal, Canada, on 5–8 May. Common proposals, debated in *Nature* and elsewhere, include improving mentorship and training, publishing negative results, reducing the pressure to publish, pre-registering studies, teaching ethics and ensuring harsh punishments.

These are important but they overestimate the benefits of correcting scientists' minds. We often forget that scientific knowledge is reliable not because scientists are more clever, objective or honest than other people, but because their claims are exposed to criticism and replication.

The key to protecting science, therefore, is to strengthen self-correction. Publication, peer-review and misconduct investigations should focus less on what scientists do, and more on what they communicate.

What is wrong with current approaches? By defining misconduct in terms of behaviours, as all countries do at present, we have to rely on whistle-blowers to discover it, unless the fabrication is so obvious as to be apparent from papers. It is rare for misconduct to have witnesses; and surveys suggest that when people do know about a colleagues' misbehaviour, they rarely report it. Investigators, then, face the arduous task of reconstructing what a scientist did, establishing that the behaviour deviated from accepted practices and determining whether such deviation expressed an intention to deceive. Only the most clear-cut cases are ever exposed.

Take the scandal of Diederik Stapel, the Dutch star psychologist who last year was revealed to have been fabricating papers for almost 20 years. How was this possible? First, Stapel insisted on collecting data by himself, which kept away potential whistle-blowers. Second, researchers had no incentive to replicate his experiments, and when they did, they lacked sufficient information to explain discrepancies. This was mainly because, third, Stapel was free to omit from papers details that would have revealed lies and statistical flaws.

In tackling these issues, a good start would be to redefine misconduct as distorted reporting: 'any omission or misrepresentation of the information necessary and sufficient to evaluate the validity and significance of research, at the level appropriate to the context in which the research is communicated'.

Some might consider this too broad. But it is no more so than the definition of falsification used by the US Office of Science and Technology Policy: "manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record". Unlike this definition, however, mine points unambiguously to misconduct whenever there is a mismatch between what was reported and what was done.

Authors should be held accountable for what they write, and for recording what they did. But who decides what information is necessary and sufficient? That would be experts in each field, who should prepare and update guidelines. This might seem daunting, but such guidelines are already being published for many biomedical techniques, thanks to initiatives such as the EQUATOR Network (equa-

tor-network.org) or Minimum Information for Biological and Biomedical Investigations (mibbi.sourceforge.net).

The main task of journal editors and referees would then be to ensure that researchers comply with reporting requirements. They would point authors to the appropriate guidelines, perhaps before the study had started, and make sure that all the requisite details were included. If authors refused or were unable to comply, their paper (or grant application or talk) would be rejected. The publication would indicate which set or sets of guidelines were followed.

By focusing on reporting practices, the community would respect scientific autonomy but impose fairness. A scientist should be free to decide, for example, that 'fishing' for statistical significance is necessary. However, guidelines would require a list of every test used, allowing others to infer the risk of false positives.

Carefully crafted guidelines could make fabrication and plagiarism more difficult, by requiring the publication of verifiable details. And they could help to uncover questionable practices such as ghost authorship, exploiting subordinates, post hoc hypotheses or dropping outliers.

Graduate students could, in addition to learning the guidelines, train by replicating published studies. Special research funds could be reserved for independent replications of unchallenged claims.

The current defence against misconduct is prepared for the wrong sort of attack: the community tries to regulate research like any other profession, but it is different. The reliability of scientific 'products' is ensured not by individual practice, but by collective dialogue. ■

**Daniele Fanelli** is a research fellow at the University of Edinburgh, UK.  
e-mail: [dfanelli@exseed.ed.ac.uk](mailto:dfanelli@exseed.ed.ac.uk)

FOCUS LESS ON WHAT  
**SCIENTISTS  
DO**  
AND MORE ON WHAT  
THEY  
**COMMUNICATE.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/jyfnuh](http://go.nature.com/jyfnuh)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## IMMUNOLOGY

### Cross-reactions boost immunity

A memory of microbes may help the immune system to fight pathogens that it has never encountered before.

Mark Davis and his colleagues at Stanford University in California examined white blood cells called T cells that carry the CD4 antigen — which quickly rouse other cells into launching a robust immune attack — in blood samples of 26 healthy adults. Although all the blood donors tested negative for HIV, the researchers found unexpectedly high levels of the T cells that recognized the virus and were primed to launch an attack. When they analysed umbilical-cord blood from newborns, however, they found that these 'memory' cells were absent, which may explain why young children are so susceptible to infection. The researchers also showed that vaccination against flu activated cells that recognize proteins not just from flu, but other bacterial species.

These findings may help to explain why vaccinations against one disease can offer protection against other infections.

*Immunity* <http://dx.doi.org/10.1016/j.immuni.2012.10.021> (2013)

## NANOTECHNOLOGY

### Lasers tune tiny diamonds

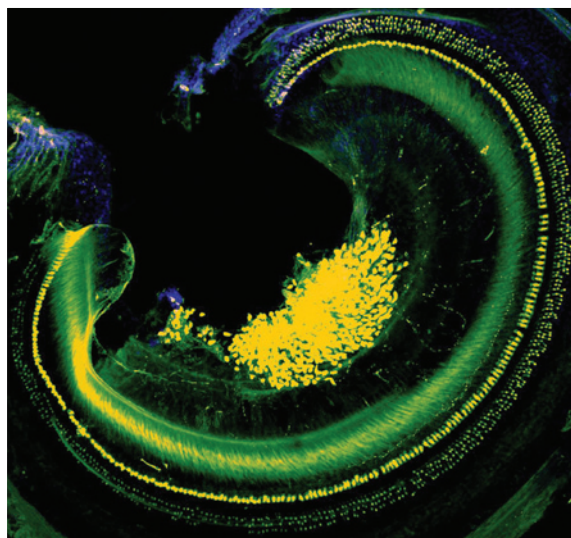
Minuscule diamond crystals show promise for applications that range from biological imaging to quantum computing, but they have been difficult to manipulate individually.

Romain Quidant of the Institute of Photonic Sciences in Barcelona, Spain, and

his colleagues developed a way to use infrared lasers to trap diamond nanocrystals with a single nitrogen atom inside. They could then change the polarization of the laser to twist and turn the nitrogen-atom axis and move individual nanodiamonds in

three dimensions. Because the technique worked for crystals suspended in solution, the researchers believe it could be used in biological systems.

*Nature Nanotechnol.* <http://dx.doi.org/10.1038/nnano.2012.259> (2013)



## MOLECULAR THERAPY

### Deafness diverted

Molecular therapy improves hearing and balance in a mouse model of a hereditary disease that causes deafness and balance disorders.

In humans, Usher syndrome is often caused by a mutation in the gene *USH1C* that results in the production of a truncated form of the protein harmonin — which normally guides development of the inner ear (pictured) — and leads to impaired hearing.

Jennifer Lentz at Louisiana State University in New Orleans, Michelle Hastings at Rosalind Franklin University in North Chicago, Illinois, and their collaborators injected newborn mice with DNA-like molecules that mask the mutation in the cell's protein-expression machinery, causing normal harmonin to be produced. This stopped the loss of inner-ear cells that respond to sound, improved low- and mid-frequency hearing and reduced behaviours associated with balance impairment such as head-tossing and circling. The effects lasted at least six months. The authors suggest that a similar approach may have therapeutic potential in human congenital deafness.

*Nature Med.* <http://dx.doi.org/10.1038/nm.3106> (2013)

## PALAEOANTHROPOLOGY

### New hominin wrist bones

Newly described wrist bones support the argument that *Homo floresiensis* is a distinct species, rather than a deformed *Homo sapiens*.

In 2003, scientists discovered parts of a skeleton (LB1) that is at least 17,000 years old in Liang Bua on the Indonesian island of Flores. The individual was described as a new hominin species, mostly based on analysis of features of its skull and lower body. The wrist bones of LB1 included features seen in many apes but not modern humans. Caley Orr, now at Midwestern University in Downers Grove, Illinois, and his colleagues have analysed wrist bones from other finds at Liang Bua and their conclusions support the original findings. The bones are smaller than those of LB1, but contain features not found in modern humans and Neanderthals.

The finding rebuts claims that the primitive features of LB1 bones were due to pathology, the authors say. *J. Hum. Evol.* 64, 109–129 (2013)

## ZOOLOGY

### Bats as disease reservoirs

Bats may be more likely than rodents to pass on a viral infection to other mammalian species, including humans, suggests a large-scale analysis.

Angela Luis of Colorado State University in Fort Collins and her colleagues searched the literature, counting and characterizing zoonotic viruses — those that can be transmitted to humans from other animals — that had been reported in bats

JENNIFER LENTZ



RAINA PLOWRIGHT

(pictured) or rodents. The authors found that bats, with an average of 1.79 viruses per species, host significantly more zoonotic viruses per species than rodents, with 1.48 viruses per species.

Bat species that live longer or produce more litters per year tended to harbour more zoonotic viruses, and a proclivity for living closely with related species was the biggest determinant of zoonotic-viral richness.

The transfer of viruses could occur more frequently between bat species, which tend to live close together, than rodent ones, and the authors suggest that this could partly explain the higher prevalence of viruses in bats. *Proc. R. Soc. B* 280, 20122753 (2013)

## LONGEVITY

## Fish oils turn on cellular recycling

The polyunsaturated fats found in fish oils may promote longevity by triggering autophagy, a process that helps cells to survive starvation conditions by degrading and recycling excess cell components.

A team led by Gary Ruvkun of Massachusetts General Hospital in Boston found that when they starved the nematode worm *Caenorhabditis elegans*, the creatures boosted their production of omega-6 polyunsaturated fatty acids — which are found in fish oils. Supplementing cells from *C. elegans* and humans with these fatty acids activated autophagy. Worms fed omega-6 fatty acids lived longer than those that did not receive the fats

— an effect erased by disabling the worms' autophagy machinery.

This mechanism could explain the health benefits that are seen in people who eat diets rich in fish oils, the researchers say.

*Genes Dev.* <http://dx.doi.org/10.1101/gad.205294.112> (2013)

## OPTOMECHANICS

## Measuring twist with light

A device that uses light to measure torsion — the amount of twist — of an object at the nanometre scale improves on the sensitivity of previous techniques.

Measuring torsion is key to studies of a wide range of forces, from gravity to electromagnetism. John Davis at the University of Alberta in Edmonton, Canada, and his colleagues have built a torsion detector consisting of a pair of paddles placed next to an optical cavity. Because the paddles' refractive index is higher than that in the cavity, their movement distorts light waves that are trapped within it. By measuring this distortion, the researchers were able to detect subtle torsional shifts in the paddle's positions with roughly 100-times greater sensitivity than other techniques that do not use light.

The researchers suggest that the technique could be useful for studying magnetic materials at the nanoscale.

*Appl. Phys. Lett.* 102, 053102 (2013)

## EVOLUTION

## Symbiosis leads to diversity

Species interactions such as competition and predation spur on diversification — as can symbiotic relationships, a study of plant-invading insects has found.

Many species of a family of

## COMMUNITY CHOICE

The most viewed papers in science

## MICROBIOLOGY

## Antibiotic strikes new target

**HIGHLY READ**  
on [aac.asm.org](http://aac.asm.org)  
in January

An antibiotic that disables protein synthesis seems to inhibit the growth of several strains of drug-resistant bacteria that are known to cause hospital infections.

Dickon Alley at Anacor Pharmaceuticals in Palo Alto, California, and his colleagues synthesized the antibacterial agent, which blocks the production of an enzyme that is essential for microbial protein synthesis and is designed to fight Gram-negative bacteria — which have an outer envelope that hampers the entry of antibiotics. The agent inhibited the growth of resistant bacteria — including the multidrug-resistant *Pseudomonas aeruginosa* — both *in vitro* and in a mouse model of infection. The properties of the antibacterial allow it to avoid the main mechanisms that Gram-negative bacteria use to fend off these drugs.

In a clinical trial, the antibiotic successfully cleared urinary-tract infections in some patients, but failed to do so in others because of bacterial resistance. The researchers say they are working to avoid this problem.

*Antimicrob. Agents Chemother.* <http://dx.doi.org/10.1128/AAC.02058-12> (2013)



JEFFREY JOY

insects known as gall midges (Cecidomyiidae) rely on fungi to help them break down plant tissues; in return, the female gall midges deposit the fungal spores along with their eggs when they move from plant to plant. In a survey of the literature, Jeffrey Joy at Simon Fraser University in Burnaby, British Columbia, Canada, found that gall midges that are associated with fungi tend to use a wider variety of host plants (pictured) compared with those with no association. Moreover, his analysis of gall-midge lineages revealed that symbiotic

insect species are more than 17 times as diverse as non-symbiotic ones.

Forming a relationship with plant-digesting fungi could allow for greater evolutionary diversity in other insect species by providing them with a greater number of potential hosts, Joy suggests.

*Proc. R. Soc. B* <http://dx.doi.org/10.1098/rspb.2012.2820> (2013)

**NATURE.COM**

For the latest research published by Nature visit:

[www.nature.com/latestresearch](http://www.nature.com/latestresearch)



# SEVEN DAYS

The news in brief

## RESEARCH

### Subglacial lake life

Researchers have found live microbes in water and sediment samples taken from Lake Whillans, a small body of water buried 800 metres beneath Antarctic ice. If the findings, confirmed to *Nature* on 10 February, hold up, it will mark the first time that life has been discovered in a subglacial lake. The US expedition team penetrated the lake late last month. See [go.nature.com/y1kcw8](http://go.nature.com/y1kcw8) for more.

### Bullying checks

Australia's national science agency, the Commonwealth Scientific and Research Organisation, says that it will appoint an external investigator to review accusations of bullying made by former employees. The agency has been troubled by complaints in recent years, and was last year told by a government watchdog to improve its systems. See [go.nature.com/yznbnz](http://go.nature.com/yznbnz) for more.

### Landsat 8 lifts off

NASA launched the United States' latest Earth-observing satellite, Landsat 8, from Vandenberg Air Force Base in California on 11 February. The satellite is due to begin

## NUMBER CRUNCH

# \$3.25 bn

The amount, plus royalties, promised by Biogen Idec, a drug company in Weston, Massachusetts, for all rights to Tysabri, a multiple-sclerosis drug developed by Elan in Dublin. Sales of Tysabri were worth US\$1.6 billion last year.



WORLD VISION/AP

## Solomon Islands tsunami kills villagers

A tsunami caused by a magnitude-8.0 quake in the South Pacific killed at least six villagers on Santa Cruz Island on 6 February, even though a timely alert from the Pacific Tsunami Warning Center in Hawaii prompted large-scale evacuations. The region — at the eastern

margin of the Australia tectonic plate — is one of Earth's most seismically active areas. Experts say that it is often difficult to communicate tsunami warnings to all parts of remote Pacific archipelagos, such as the Solomon Islands. See [go.nature.com/z6zjk3](http://go.nature.com/z6zjk3) for more.

operations within 90 days, and will extend and improve a 40-year data record of global environmental change, helping scientists to track and study issues including deforestation and melting glaciers. See [go.nature.com/wnykfa](http://go.nature.com/wnykfa) for more.

## POLICY

### Fishing reforms

The European Parliament approved a major shake-up of the region's fishing industry in a vote on 6 February. Scientists have long called for changes to the European Union's Common Fisheries Policy, which they blame for some species being fished at unsustainable levels. The parliament voted to end the practice of throwing back fish

that are not the target species or are too small, which is wasteful because these fish often die anyway. The vote also supported improved practices for managing fish populations to maintain the maximum possible harvest on a permanent basis. The proposals will be finalized in discussions between the parliament, the European Council and the European Commission. See [go.nature.com/2swipm](http://go.nature.com/2swipm) for more.

### Chemical safety

The world's most extensive chemical-safety regulation, implemented five years ago in Europe, is working well, says a review published by the European Commission on 5 February. The REACH regulation (registration,

evaluation, authorization and restriction of chemicals) requires companies that produce or sell chemicals in Europe to register toxicity data on the compounds and to propose tests needed to clarify their biological effects. The law has improved the quality and quantity of information available about chemicals on the market, says the review, but European Union member states need to make more effort to ensure that companies comply. See [go.nature.com/chuxxs](http://go.nature.com/chuxxs) for more.

### EU budget deal

European leaders agreed on 8 February to a €125.6-billion (US\$168-billion) budget for the period 2014–20 for initiatives to boost Europe's competitiveness, including

the Horizon 2020 research programme. See pages 147 and 159 for more.

## PEOPLE

## Minister resigns

Germany's science and education minister, Annette Schavan, resigned on 9 February, four days after the University of Düsseldorf confirmed accusations of plagiarism in her doctoral thesis and revoked her PhD. German Chancellor Angela Merkel has nominated Johanna Wanka (former education minister in the state of Lower Saxony) to succeed Schavan. On 5 February, a university inquiry concluded that Schavan had "systematically and deliberately" reproduced work that was not her own without proper reference in her 1980 philosophy thesis. Schavan said that she would challenge that decision in court. See [go.nature.com/9gcstt](http://go.nature.com/9gcstt) for more.

## NSF director quits

Subra Suresh, director of the US National Science Foundation (NSF), announced his resignation on 5 February, cutting short a six-year term after less than three years. He will leave the basic-research agency in late March to become president of Carnegie Mellon University in Pittsburgh, Pennsylvania.



Suresh (pictured) earned a reputation at the NSF for promoting interdisciplinary research and international collaborations. He leaves the foundation at a critical time: all government agencies face across-the-board budget cuts on 1 March unless lawmakers resolve a political deadlock over the federal budget. See [go.nature.com/noxpd](http://go.nature.com/noxpd) and page 158 for more.

## Interior secretary

US President Barack Obama nominated Sally Jewell, the chief executive of a major outdoor-recreation company, to succeed Ken Salazar as secretary of the interior on 6 February. Jewell currently heads Recreational Equipment Inc. (REI), based in Kent, Washington, and is a former petroleum engineer and banker. She is also a conservation advocate and outdoor enthusiast. If her nomination is agreed by the

Senate, Jewell will oversee decisions about hydraulic fracturing and renewable-energy permits on public lands.

## Wellcome stand-in

One of the world's biggest biomedical funders, the Wellcome Trust, announced on 6 February that Ted Bianco will become its acting director while the charity searches for a permanent replacement for Mark Walport. Walport will leave his post as director on 14 March to become the British government's next chief science adviser. Bianco is currently the charity's director of technology transfer. See [go.nature.com/vqmhl1](http://go.nature.com/vqmhl1) for more.

## BUSINESS

## Open-access launch

A new open-access journal, *PeerJ*, which offers authors a lifetime of free publication for a one-off membership fee of US\$299, published its first batch of articles on 12 February. Keenly watched for its unusual business model, the journal could dramatically drive down the costs of publishing, its founders say (see *Nature* 486, 166; 2012).

## EVENTS

## Polio killings

Gunmen burst into two clinics in the northern Nigerian city of Kano on 8 February and

## COMING UP

### 14–18 FEBRUARY

The American Association for the Advancement of Science holds its annual meeting in Boston, Massachusetts.

[www.aaas.org/meetings](http://www.aaas.org/meetings)

### 19 FEBRUARY

The US Supreme Court hears arguments in *Bowman v. Monsanto*, a case to determine whether patent protections extend to the subsequent generations of seeds derived from genetically engineered plants.

[go.nature.com/vbrack](http://go.nature.com/vbrack)

killed at least nine people, including polio-vaccination workers. No one has claimed responsibility, but the militant Islamist group Boko Haram is widely suspected. Last year, Nigeria recorded more cases of polio than any other country in the world, but no new cases of the paralyzing disease have been reported since December.

## Nuclear test

North Korea conducted a third underground nuclear-weapons test on 12 February. The test was detected by seismic monitoring stations of the US Geological Survey and the Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) in Vienna, which reported "explosion-like characteristics". The yield is believed to be several kilotonnes, and roughly twice the size of the nation's last nuclear test in 2009, according to the CTBTO. It comes two months after North Korea launched its first satellite into orbit. See [go.nature.com/dbrsav](http://go.nature.com/dbrsav) for more.

➔ [NATURE.COM](http://NATURE.COM)

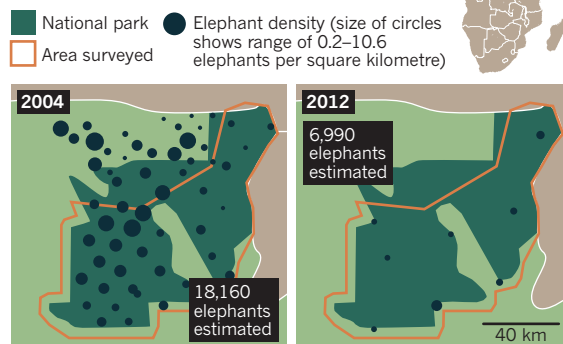
For daily news updates see:  
[www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

Ivory poachers killed around 11,000 elephants in Minkébé National Park in Gabon between 2004 and 2012, according to figures published on 6 February. Gabon is home to more than half of Africa's forest elephants, and most of those roam the Minkébé park. A study by the Gabonese National Parks Agency and conservation organizations the WWF and the Wildlife Conservation Society says that populations are now so reduced that the species could disappear locally.

## AFRICAN ELEPHANT NUMBERS SLASHED

Poaching has cut the largest population of forest elephants in Africa by three-fifths in the past eight years.



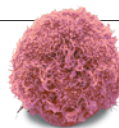


# NEWS IN FOCUS

**GENETIC DIAGNOSIS** Data barriers hamper search for meaning in mutations **p.156**

**FUNDING** US science agencies gird themselves for the budget axe **p.158**

**MALARIA** Plant source of key drug faces lab-made competition **p.160**



**BIOMEDICINE** A Texas-style showdown over stem-cell therapy **p.166**

JOHN ANGELL/LOUP/NEWS.COM



The latest US influenza season is more severe and has caused more deaths than usual.

## EPIDEMIOLOGY

# When Google got flu wrong

*US outbreak foxes a leading web-based method for tracking seasonal flu.*

BY DECLAN BUTLER

When influenza hit early and hard in the United States this year, it quietly claimed an unacknowledged victim: one of the cutting-edge techniques being used to monitor the outbreak. A comparison with traditional surveillance data showed that Google Flu Trends, which estimates prevalence from flu-related Internet searches, had drastically overestimated peak flu levels. The glitch is no more than a temporary setback for a promising strategy, experts say, and Google is sure to refine its algorithms. But as flu-tracking techniques based on mining of web data and on social media proliferate, the episode is a reminder that they will

complement, but not substitute for, traditional epidemiological surveillance networks.

"It is hard to think today that one can provide disease surveillance without existing systems," says Alain-Jacques Valleron, an epidemiologist at the Pierre and Marie Curie University in Paris, and founder of France's Sentinelles monitoring network. "The new systems depend too much on old existing ones to be able to live without them," he adds.

This year's US flu season started around November and seems to have peaked just after Christmas, making it the earliest flu season since 2003. It is also causing more serious illness and deaths than usual, particularly among the elderly, because, just as in 2003, the predominant strain this year is H3N2 — the most

virulent of the three main seasonal flu strains.

Traditional flu monitoring depends in part on national networks of physicians who report cases of patients with influenza-like illness (ILI) — a diffuse set of symptoms, including high fever, that is used as a proxy for flu. That estimate is then refined by testing a subset of people with these symptoms to determine how many have flu and not some other infection.

With its creation of the Sentinelles network in 1984, France was the first country to computerize its surveillance. Many countries have since developed similar networks — the US system, overseen by the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, includes some 2,700 health-care centres that record about 30 million patient visits annually.

But the near-global coverage of the Internet and burgeoning social-media platforms such as Twitter have raised hopes that these technologies could open the way to easier, faster estimates of ILI, spanning larger populations.

The mother of these new systems is Google's, launched in 2008. Based on research by Google and the CDC, it relies on data mining records of flu-related search terms entered in Google's search engine, combined with computer modelling. Its estimates have almost exactly matched the CDC's own surveillance data over time — and it delivers them several days faster than the CDC can. The system has since been rolled out to 29 countries worldwide, and has been extended to include surveillance for a second disease, dengue.

Google Flu Trends has continued to perform remarkably well, and researchers in many countries have confirmed that its ILI estimates are accurate. But the latest US flu season seems to have confounded its algorithms. Its estimate for the Christmas national peak of flu is almost double the CDC's (see 'Fever peaks'), and some of its state data show even larger discrepancies.

It is not the first time that a flu season has tripped Google up. In 2009, Flu Trends had to tweak its algorithms after its models badly underestimated ILI in the United States at the start of the H1N1 (swine flu) pandemic — a glitch attributed to changes in people's search

behaviour as a result of the exceptional nature of the pandemic (see <http://doi.org/djw73f>).

Google would not comment on this year's

**NATURE.COM**  
See maps showing reports of flu-like symptoms in France: [go.nature.com/w954hn](http://go.nature.com/w954hn)



► difficulties. But several researchers suggest that the problems may be due to widespread media coverage of this year's severe US flu season, including the declaration of a public-health emergency by New York state last month. The press reports may have triggered many flu-related searches by people who were not ill. Few doubt that Google Flu will bounce back after its models are refined, however.

"You need to be constantly adapting these models, they don't work in a vacuum," says John Brownstein, an epidemiologist at Harvard Medical School in Boston, Massachusetts. "You need to recalibrate them every year."

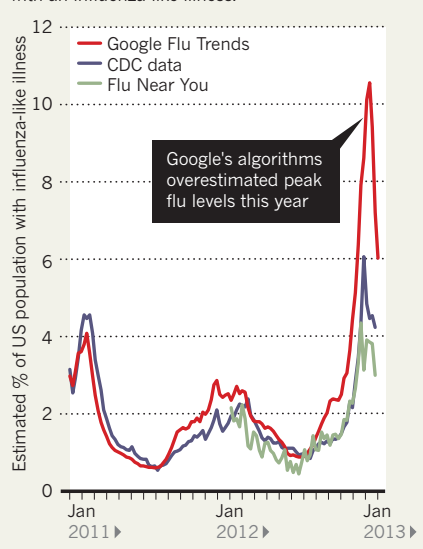
Brownstein is one of many researchers trying to harness the power of the web to establish sentinel networks made up not of physicians, but of ordinary citizens who volunteer to report when they or someone in their family are experiencing symptoms of ILI. 'Flu Near You', a system run by the HealthMap initiative co-founded by Brownstein at Boston Children's Hospital, was launched in 2011 and now has 46,000 participants, covering 70,000 people.

Similar systems are springing up in Europe. For example, GrippeNet.fr, run by French researchers in collaboration with national health authorities, has attracted more than 5,500 participants since its creation a year ago, with 60–90 people joining each week.

Lyn Finelli, head of the CDC's Influenza Surveillance and Outbreak Response Team, feels that such crowdsourcing techniques hold great promise, especially because the questionnaires are based on clinical definitions of ILI and so yield very clean data. And both Flu Near You and GrippeNet.fr have a representative age distribution of participants. The CDC has

## FEVER PEAKS

A comparison of three different methods of measuring the proportion of the US population with an influenza-like illness.



worked with Flu Near You on its development, and Finelli herself has signed up: "I submit my family's data every week," she says.

Other researchers are turning to what is probably the largest publicly accessible alternative trove of social-media data: Twitter. Several groups have published work suggesting that models of flu-related tweets can be closely fitted to past official ILI data, and various services, such as MappyHealth and Sickweather, are testing whether real-time analyses of tweets can reliably assess levels of flu.

But Finelli is sceptical. "The Twitter analyses

have much less promise" than Google Flu or Flu Near You, she says, arguing that Twitter's signal-to-noise ratio is very low, and that the most active Twitter users are young adults and so are not representative of the general public.

Michael Paul, a computer scientist at Johns Hopkins University in Baltimore, Maryland, disagrees. He is part of a team that is developing Twitter-based disease monitoring, and says that Google search-term data probably have just as much noise. And although Internet-based surveys may boast less noise, their smaller size means that they may be prone to sampling errors. "I suspect that passive monitoring of social media will always yield more data than systems that rely on people to actively respond to surveys, like Flu Near You," Paul says.

To reduce the noise, the Johns Hopkins team has recently analysed a subset of a few thousand flu-related tweets, looking for patterns indicating which tweets showed that the tweeter was actually ill rather than simply, say, pointing to news articles about flu. They then used this information to retrain their models to weed out irrelevant flu-related tweets. Paul says that a paper in press will show that this greatly improves their results.

Already, web data mining and crowdsourced tracking systems are becoming a part of the flu-surveillance landscape. "I'm in charge of flu surveillance in the United States and I look at Google Flu Trends and Flu Near You all the time, in addition to looking at US-supported surveillance systems," says Finelli. "I want to see what's happening and if there is something that we are missing, or whether there is a signal represented somewhat differently in one of these other systems that I could learn from." ■

SOURCES: GOOGLE FLU TRENDS (WWW.GOOGLE.ORG/FLUTRENDS); CDC; FLU NEAR YOU

## MEDICINE

# Data barriers limit genetic diagnosis

*Tools for data-sharing promise to improve chances of connecting mutations with symptoms of rare diseases.*

BY ERIKA CHECK HAYDEN

For the first five months of Harrison Harkins' life, doctors had little idea about what was causing his spinal malformation and inability to gain weight. But in November 2011, Matthew Bainbridge, a computational biologist at Baylor College of Medicine in Houston, Texas, found a clue. After analysing genetic data from Harrison and his parents, Bainbridge discovered that the child had an

abnormal version of a gene called *ASXL3*.

But Bainbridge had no easy access to records of other children with *ASXL3* mutations, and could not be sure that this mutation was the culprit. So he did what many scientists do: he networked. A Dutch team put Bainbridge in touch with German researchers who were treating another boy with an *ASXL3* mutation — and symptoms similar to Harrison's. After finding two further cases in an internal Baylor database, Bainbridge felt that the

connection was concrete. He describes the syndrome seen in all four children, and probably caused by *ASXL3* mutations, in a paper published on 5 February (M. N. Bainbridge *et al. Genome Med.* 5, 11; 2013).

Researchers are using new tools to increase the pace of discoveries such as Bainbridge's. Efforts to connect sequences with symptoms — or in genetic parlance, genotype with phenotype — have taken on increased urgency as clinical sequencing gains traction and funders put more money towards rare diseases. Researchers are planning to address the barriers to data sharing at a workshop in April, after the first International Rare Diseases Research Consortium Conference in Dublin. "There is a very positive feeling in the community that things are changing for the better," says Peter Robinson, a computational biologist at the Charity University Hospital in Berlin.

Thousands of people have had their genomes sequenced, but a reluctance to surrender ownership of the valuable data, along with the privacy concerns of researchers and families (see 'Families find solace in

► difficulties. But several researchers suggest that the problems may be due to widespread media coverage of this year's severe US flu season, including the declaration of a public-health emergency by New York state last month. The press reports may have triggered many flu-related searches by people who were not ill. Few doubt that Google Flu will bounce back after its models are refined, however.

"You need to be constantly adapting these models, they don't work in a vacuum," says John Brownstein, an epidemiologist at Harvard Medical School in Boston, Massachusetts. "You need to recalibrate them every year."

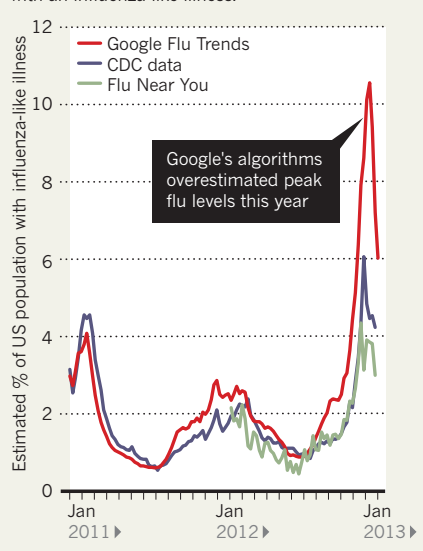
Brownstein is one of many researchers trying to harness the power of the web to establish sentinel networks made up not of physicians, but of ordinary citizens who volunteer to report when they or someone in their family are experiencing symptoms of ILI. 'Flu Near You', a system run by the HealthMap initiative co-founded by Brownstein at Boston Children's Hospital, was launched in 2011 and now has 46,000 participants, covering 70,000 people.

Similar systems are springing up in Europe. For example, GrippeNet.fr, run by French researchers in collaboration with national health authorities, has attracted more than 5,500 participants since its creation a year ago, with 60–90 people joining each week.

Lyn Finelli, head of the CDC's Influenza Surveillance and Outbreak Response Team, feels that such crowdsourcing techniques hold great promise, especially because the questionnaires are based on clinical definitions of ILI and so yield very clean data. And both Flu Near You and GrippeNet.fr have a representative age distribution of participants. The CDC has

## FEVER PEAKS

A comparison of three different methods of measuring the proportion of the US population with an influenza-like illness.



worked with Flu Near You on its development, and Finelli herself has signed up: "I submit my family's data every week," she says.

Other researchers are turning to what is probably the largest publicly accessible alternative trove of social-media data: Twitter. Several groups have published work suggesting that models of flu-related tweets can be closely fitted to past official ILI data, and various services, such as MappyHealth and Sickweather, are testing whether real-time analyses of tweets can reliably assess levels of flu.

But Finelli is sceptical. "The Twitter analyses

have much less promise" than Google Flu or Flu Near You, she says, arguing that Twitter's signal-to-noise ratio is very low, and that the most active Twitter users are young adults and so are not representative of the general public.

Michael Paul, a computer scientist at Johns Hopkins University in Baltimore, Maryland, disagrees. He is part of a team that is developing Twitter-based disease monitoring, and says that Google search-term data probably have just as much noise. And although Internet-based surveys may boast less noise, their smaller size means that they may be prone to sampling errors. "I suspect that passive monitoring of social media will always yield more data than systems that rely on people to actively respond to surveys, like Flu Near You," Paul says.

To reduce the noise, the Johns Hopkins team has recently analysed a subset of a few thousand flu-related tweets, looking for patterns indicating which tweets showed that the tweeter was actually ill rather than simply, say, pointing to news articles about flu. They then used this information to retrain their models to weed out irrelevant flu-related tweets. Paul says that a paper in press will show that this greatly improves their results.

Already, web data mining and crowdsourced tracking systems are becoming a part of the flu-surveillance landscape. "I'm in charge of flu surveillance in the United States and I look at Google Flu Trends and Flu Near You all the time, in addition to looking at US-supported surveillance systems," says Finelli. "I want to see what's happening and if there is something that we are missing, or whether there is a signal represented somewhat differently in one of these other systems that I could learn from." ■

SOURCES: GOOGLE FLU TRENDS (WWW.GOOGLE.ORG/FLUTRENDS); CDC; FLU NEAR YOU

## MEDICINE

# Data barriers limit genetic diagnosis

*Tools for data-sharing promise to improve chances of connecting mutations with symptoms of rare diseases.*

BY ERIKA CHECK HAYDEN

For the first five months of Harrison Harkins' life, doctors had little idea about what was causing his spinal malformation and inability to gain weight. But in November 2011, Matthew Bainbridge, a computational biologist at Baylor College of Medicine in Houston, Texas, found a clue. After analysing genetic data from Harrison and his parents, Bainbridge discovered that the child had an

abnormal version of a gene called *ASXL3*.

But Bainbridge had no easy access to records of other children with *ASXL3* mutations, and could not be sure that this mutation was the culprit. So he did what many scientists do: he networked. A Dutch team put Bainbridge in touch with German researchers who were treating another boy with an *ASXL3* mutation — and symptoms similar to Harrison's. After finding two further cases in an internal Baylor database, Bainbridge felt that the

connection was concrete. He describes the syndrome seen in all four children, and probably caused by *ASXL3* mutations, in a paper published on 5 February (M. N. Bainbridge *et al. Genome Med.* 5, 11; 2013).

Researchers are using new tools to increase the pace of discoveries such as Bainbridge's. Efforts to connect sequences with symptoms — or in genetic parlance, genotype with phenotype — have taken on increased urgency as clinical sequencing gains traction and funders put more money towards rare diseases. Researchers are planning to address the barriers to data sharing at a workshop in April, after the first International Rare Diseases Research Consortium Conference in Dublin. "There is a very positive feeling in the community that things are changing for the better," says Peter Robinson, a computational biologist at the Charity University Hospital in Berlin.

Thousands of people have had their genomes sequenced, but a reluctance to surrender ownership of the valuable data, along with the privacy concerns of researchers and families (see 'Families find solace in

## A RARE CONNECTION

## Families find solace in sequencing

BRIAN MCGUCKIN

Medical researchers versed in the power of genetic sequencing may be surprised by how unaware of it — or doubtful about its benefits — the rest of the world remains. Tim Harkins, who develops sequencing technologies at Life Technologies in Carlsbad, California, found this out when he suggested that doctors sequence part of his ailing son's genome to help with the child's diagnosis. They "didn't have a clue what I was talking about", says Harkins.

So he approached geneticists at Baylor College of Medicine in Houston, Texas, who discovered a mutation in Harrison's ASXL3 gene. The diagnosis "was an invaluable source of comfort", says Harkins, not least because he could tell his older son, now aged 13, that the mutation was spontaneous and not heritable. "I was able to tell him, 'You have a better chance of being struck by lightning than having a baby like Harrison,'" says Harkins.

Despite the rapidly dropping cost of sequencing, Harkins notes, there is still a perception that it is too expensive for clinical



Harrison Harkins died one month after his final diagnosis.

use and does not deliver meaningful results. But his own experience gives the lie to that sentiment, he says. "Knowing the diagnosis can prevent a lot of needless surgeries and treatments and it eliminates the guessing, the 'Why don't we try this or that?'"

The unease runs even deeper in Germany. Maria and Christian-Alexander Neuling, who live in Hamburg, found some comfort after a genetic diagnosis showed that their son, Ferdinand, also had an ASXL3 mutation. The diagnosis helped to reassure the Neulings that they were doing everything they could for Ferdinand, who is now four; it also enabled them to connect with another family in the United States whose child has the same condition.

But they worry that genetic sequencing is viewed too sceptically in Germany, where a doctor's orders and mandatory counselling are required for any genetic test. "The discussion here is pretty one-sided; there is a strong lobby opposing genetic sequencing," says Christian-Alexander Neuling. Neuling and his wife hope that their story will assuage some of the fears of other Germans, who worry that sequencing will lead to eugenics and 'designer babies'. "A parent will go from hell to heaven and back to find something that will help their sick child," he says. **E.C.H.**

sequencing') often keep scientists from comparing findings. Many data are also off-limits because they are held by private diagnostic companies. "It's a big conundrum for labs that are doing sequencing for diagnostic services," says Michael Bamshad, chief of paediatric genetic medicine at the University of Washington in Seattle. "If they find a variant in a gene, how do they know the variant is causal?"

Patients with rare, difficult-to-diagnose disorders stand to gain the most from increased data sharing. Scientists have found the genetic roots of fewer than half of the 7,000 known rare heritable diseases, but a diagnosis can give parents an idea of a child's outlook, and give researchers a target for drug development.

Several groups are trying to build richer databases and get them to communicate. In November, for instance, the US National Center for Biotechnology Information in

Bethesda, Maryland, set up a database called ClinVar, which pools information from dozens of other databases, and allows labs to deposit data on mutations seen in individual patients (see *Nature* **491**, 171; 2012).

Still, says Anthony Brookes, a geneticist at the University of Leicester, UK, many diagnostic labs are unable to share information with databases such as ClinVar, either because they do not have the time or the expertise in depositing data, or because they are afraid that they might compromise patient security and their own livelihoods. "It's not their role to put data out there for researchers to play with," he says.

Brookes is trying to address this problem with a tool called Cafe Variome, which he describes as more of a "shop window" than a database. Labs submit information about what data they have to Cafe Variome. Users can then browse the website to see what data exist, and,

if interested, can follow up with the relevant labs. That allows the labs to control who sees their data, and to be credited when it is used. They are "much more comfortable sharing if they know the data are only being accessed by other diagnostic labs", says Brookes.

Another problem is that even if database owners are willing to share data, they lack a common language for describing phenotypes, says Robinson. He is working on ways to standardize phenotype definitions for large-scale analysis.

For researchers such as Bainbridge, the tools can't come quickly enough. His team's final diagnosis came just a month before Harrison died last March, at the age of 9 months. "If you spent 15 minutes with the parents of any of these children, you would know that everyone should be doing this," says Bainbridge. "This is going to help a lot of people at really low cost." ■

  
**MORE  
ONLINE**

## TOP STORY

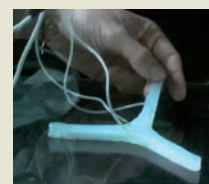


Data reveal common ancestor of placental mammals  
[go.nature.com/9ezari](http://go.nature.com/9ezari)

## MORE NEWS

- Plant experiments renew controversy on extra-genome inheritance [go.nature.com/ukvsc3](http://go.nature.com/ukvsc3)
- Small molecule triggers cancer-cell suicide in mice [go.nature.com/riagfg](http://go.nature.com/riagfg)
- How to turn living cells into computers [go.nature.com/1v6bzb](http://go.nature.com/1v6bzb)

## VIDEO



Controlled explosions make rubbery robot jump  
[go.nature.com/18395v](http://go.nature.com/18395v)

CARL BUELL





Biophysicist Daniel Kiehart's lab is in limbo until he hears about a grant renewal.

## POLICY

# Science agencies prepare for cuts

*Scientists already feeling the bite of US budget sequester.*

BY MEREDITH WADMAN

Daniel Kiehart, a biophysicist at Duke University in Durham, North Carolina, has put off the US\$8,000 recharge of the gas laser he uses to probe a key motor protein in *Drosophila* embryos. In December, he bought his own airline ticket to a meeting in San Francisco, California. And two weeks ago, when his senior postdoc handed in her notice, he did not move to replace her. "I just can't do that right now," says Kiehart.

In normal times, the US National Institutes of Health (NIH) in Bethesda, Maryland, would have already renewed a grant that Kiehart has held since 1984; his proposal was rated between "excellent" and "outstanding" last summer. Yet these are anything but normal times, and Kiehart is still awaiting a decision.

On 1 March, agencies such as the NIH and the National Science Foundation (NSF), based in Arlington, Virginia, face an abrupt 5.1% cut in this year's spending under the 'sequester' — a cut that will be all the more painful because it must be done before the US fiscal year ends on 30 September (see 'US budget woes'). Although Congress might reach a last-minute deal to delay or avoid some of the reductions, agencies are wary of committing themselves to grants that they might not be able to afford, and scientists are starting to feel the sting.

The agencies are "making very conservative decisions because nobody wants to overspend and be caught", says Howard Garrison, deputy executive director for policy at the Federation of American Societies for Experimental Biology in Bethesda, Maryland.

The agencies will have some flexibility in apportioning the cuts, which amount to \$1.57 billion at the NIH and \$288 million at the NSF. At the NIH, the 5.1% reduction would be applied to each of the 27 institutes and centres — with some protection for its Clinical Center, the research hospital in Bethesda where cuts could put patients' lives at stake. Institute directors could cut some programmes more heavily than others, as long as the total reduction equals 5.1% (see 'Wiggle room').

At the NSF, it is expected that the cuts will be applied to each major funding account. One of these, the research account, includes scientific directorates from geosciences and mathematics to engineering and biological sciences, and comprises \$5.7 billion of the agency's \$7-billion budget. Like the NIH institute directors, departing NSF director Subra Suresh or his successor would be able to protect some research programmes from the brunt of the cuts. Since 2010, NSF research budgets have favoured computing and engineering. "I suspect he is going to try to protect his priorities as much as possible," says Joel Widder, a former NSF deputy director

who is now a partner at the Oldaker Group in Washington DC, a lobbying group for universities and research organizations.

Senior officials at the science agencies, under orders from the White House Office of Management and Budget (OMB), will not discuss their plans for the possible cuts. But an OMB memo last month directed each agency to minimize impacts on its core mission, and to anticipate challenges that could raise concerns over life, safety or health. Although some of these are obvious — scientists at the isolated stations of the NSF-funded Antarctic research programme need to be equipped for the winter, for instance — others are less so. "People have not been given sufficient time to do anything but take a meat axe to the portfolio," notes Garrison.

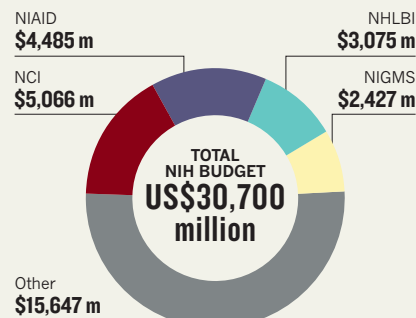
At the NIH, some of the vulnerabilities are already clear. For example, the agency's \$3.4-billion intramural research programme will be squeezed. Because much of that budget is tied up in the salaries of its more than 18,000 employees (whose jobs are expected to remain secure), other spending, from mouse cages to pipette orders, would have to absorb the losses.

There is another pressure point: some \$16 billion of the agency's \$30.7-billion total budget is tied up in multi-year grant commitments. Without significant cuts to these, the rest of the NIH budget would have to be slashed further. Already, the agency is paying instalments on previously awarded grants at only 90% of what was promised — a typical strategy when Congress has not agreed on final budget numbers. But grant recipients normally get the remaining money, or most of it, late in the fiscal year. In the event of a sequester, grant recipients would be highly unlikely to see more than half of the remaining 10% delivered. That could leave principal investigators deliberating between buying lab animals or firing postdocs.

For those seeking new funding or grant renewals, the choices may be even more grim. In a normal February, Ellen Ketterson, an evolutionary ecologist and long-time NSF grant recipient at Indiana University in

## WIGGLE ROOM

The National Institutes of Health would apply 5.1% cuts at the level of its 27 institutes and centres — giving directors the flexibility to pick favourites within those institutes.



NIAID, National Institute of Allergy and Infectious Diseases; NHLBI, National Heart, Lung and Blood Institute; NCI, National Cancer Institute; NIGMS, National Institute of General Medical Sciences.

SOURCE: NIH

## US BUDGET WOES

*Spending battles loom in coming weeks*

The trap that the US Congress devised two years ago to force cuts in the federal deficit is about to spring. Steep budget cuts, known as sequestration, are scheduled to hit on 1 March. In January, a deadlocked Congress delayed the cuts at the last minute (see *Nature* **493**, 13; 2013), but this time, many Capitol Hill observers say that there is little appetite for compromise. “I have every reason to believe that sequestration is going to go into effect,” says Chris Hellman of the National Priorities Project, an organization in Northampton, Massachusetts, that promotes budget transparency.

Scientists are already feeling the effects of the impending cuts (see main story),

expected to amount to 5.1% at the National Science Foundation and the National Institutes of Health. But because all of the cuts for 2012 must be squeezed into the 7 months before the fiscal year ends on 30 September, they “will feel like closer to 10%”, says Barry Toiv, a spokesman for the Association of American Universities in Washington DC.

The delayed 2013 budget appropriations are unlikely to alleviate the pain. On 27 March, the stop-gap spending bill that has sustained government agencies at 2012 levels will expire. If cuts come on 1 March, science advocates may not have time to fight to restore funding in any 2013 budget agreement. “I just don’t see us being able to

get that back,” says Jennifer Zeitzer, director of legislative affairs for the Federation of American Societies for Experimental Biology in Bethesda, Maryland. “That just means we’ll be asking for more in 2014.”

The budgetary logjam has also delayed President Barack Obama’s budget request for 2014. The proposal was due in early February, but is now expected to debut in March. Experts say that it is difficult to predict what the president will propose for science agencies, or whether Congress will be receptive. “It’s going to be a big fight,” says Michael Lubell, director of public affairs for the American Physical Society in Washington DC. **Helen Shen**

Bloomington, would be hiring three field assistants and ordering mist nets, traps and colour bands for a census of dark-eyed junco songbirds that she has conducted in Virginia each spring since 1985. Ketterson says that continuity is crucial to understanding how long the

birds live, when they migrate and how their breeding dates are affected by climate change.

She applied for an NSF grant last August, and expected to learn of her award in November. But her programme officer told her that the application is on hold, with no chance of

funding before the NSF learns what its budget will be. She is approaching former field staff to see whether they can donate their time, and she is applying for emergency bridge funding from the university. “There are others, also waiting,” she says. “Delay has consequences.” ■

## FUNDING

# Europe scales back research plans

*Leaders propose 13% cut to commission’s proposals.*

BY ALISON ABBOTT

**T**he ambitious vision for Europe’s next research programme dimmed last week when heads of the 27 member states of the European Union (EU) agreed to slash the overall budget for 2014–20, at the end of a long and acrimonious meeting. The deal scales back the budget proposed by the European Commission in November 2011 for its Horizon 2020 research programme by about 13%, to €69.24 billion (US\$108 billion), and means that the first year of the new programme will have a smaller budget than the last year of the previous one.

“The way it looks now, with all the arguments that were on the table in favour of promoting research in Europe, that decision is disappointing,” says Helga Nowotny, president of the European Research Council (ERC).

Europe’s leaders also reduced funding for specific scientific projects, including Galileo,

Europe’s satellite-navigation system, but they urged poor regions and countries to use more of their EU subsidies to pay for science.

A spokesman for the commission points out that even the reduced budget for Horizon 2020 marks a significant rise in research funding compared with the €55 billion for research during 2007–13.

The highly regarded ERC had been counting on a larger increase. Founded in 2007, it awards large grants on the basis of research excellence, and had hoped for a €13-billion slice of Horizon 2020 to help boost the current 12% success rate of proposals. The council did not detail how the cuts should be distributed among Horizon 2020’s research programmes.

The European Parliament has yet to endorse the council’s decision; politicians there will cast their votes some time in the next three months. Parliament had called for a mighty €100-billion budget for Horizon 2020 (see *Nature* **489**, 188–189; 2012).

Some prominent parliamentarians, including Christian Ehler, rapporteur for Horizon 2020, have vowed to fight for more. But observers say that behind-the-scenes negotiations between the commission, council and parliament over the past months suggest that the parliament is unlikely to get its own way.

The council’s proposal that more EU cohesion funds — historically used to promote the competitiveness of poor regions by improving infrastructure such as transport links — be used for research has precedent. Last year, Greece awarded €3.7 million in cohesion funds to molecular biologist George Kollias from the Biomedical Sciences Research Centre ‘Alexander Fleming’ near Athens. He will use them to equip and operate a Greek node of the Infrafrontier project, which aims to systematically phenotype, or characterize, mutant mice. “It’ll give all Greek scientists easier access to high-tech genetic tools,” he says.

The council specified exact budgets for three large scientific infrastructure programmes outside Horizon 2020. Its recommendations cut the commission’s budget for Galileo by 10% and for GMES, a system of Earth-observation satellites, by roughly one-third. The council did, however, allocate €2.7 billion to the experimental nuclear fusion reactor ITER, which the commission had proposed not funding.

“But discussions are not yet over,” cautions Nowotny. “We all have to wait to see the final figures and agreement.” ■ **SEE EDITORIAL P.147**



# Malaria drug made in yeast causes market ferment

*Synthetic biology delivers combination therapies into an uncertain market.*

BY MARK PELOW

“It’s been a dream project — but it’s been a long dream,” says Jay Keasling, a biochemical engineer at the University of California, Berkeley. Seven years ago, he and his team genetically engineered yeast to produce artemisinic acid (D.-K. Ro *et al.* *Nature* **440**, 940–943; 2006), a precursor to the best malaria treatments available: artemisinin-based combination therapies (ACTs). Synthetic biology, Keasling hoped, could produce the drug more cheaply and reliably than natural sources, benefiting the roughly 200 million people infected with malaria each year.

Keasling’s pipe dream has turned into a drug pipeline. In 2008, Paris-based pharmaceutical company Sanofi licensed the yeast that he helped to develop, and at an artemisinin conference in Nairobi last month, Keasling learned that the company had produced almost 39 tonnes of artemisinic acid — the first industrial-scale deployment of synthetic biology for drug production. The stock could be converted to at least 40 million treatments, says Keasling.

But the elegant science faces some messy realities. This year will see the end of one of the main funding routes for ACTs — the Affordable Medicines Facility — Malaria (AMFm) programme, run by the Global Fund to Fight AIDS, Tuberculosis and Malaria in Geneva, Switzerland. Its demise may not leave enough alternative funding to pay for the extra treatments made possible by the semi-synthetic process. Furthermore, if Sanofi’s product is rushed into pharmacies at similar prices to existing products, it could disrupt an already volatile market (see *Nature* **466**, 672–673; 2010). “If it’s brought in too fast it could create huge shortages, because people will stop producing the natural stuff,” says Malcolm Cutler, technical adviser to the Assured Artemisinin Supply System initiative, which organized the Nairobi conference.

The sweet wormwood plant, *Artemisia annua*, has until now been the only source of artemisinin. Drug-makers convert the natural product into derivatives that are more easily taken up by the body, and then combine them with other drugs to prevent the malaria parasite from

➔ **NATURE.COM**  
For more, see a  
Nature Outlook on  
malaria.  
[go.nature.com/spwwfg](http://go.nature.com/spwwfg)



LAURA NEWMAN/PATH

People in malaria-prone countries could soon be treated with drugs made by engineered yeast.

developing resistance. Since 2005, when ACTs became the preferred treatment of the World Health Organization (WHO), governments and health policy-makers have encouraged farmers in China, Vietnam and elsewhere to grow *A. annua* to meet rising demand. But the dependence on erratic donor funding, along with a production cycle that can take a year and a half from planting to drug production, has made it difficult to manage that supply. Last year, a bumper crop of *A. annua* and fears of declining funding sent artemisinin prices plummeting (see ‘The cost of progress’).

Yeast make ACT production faster and more

manageable. The microbes are not affected by weather, and they churn out artemisinic acid that is readily turned into semi-synthetic artemisinin, which Sanofi or other makers can convert into ACT ingredients such as artesunate. The whole process takes less than three months. “You get an order, you fire up the bioreactors, and you ship it out,” says Keasling. Sanofi hopes to get WHO approval for semi-synthetic artemisinin in the coming months, says Ponni Subbiah, a drug-development programme leader at PATH, a global health organization based in Seattle, Washington, which has funded and coordinated the development of semi-synthetic artemisinin.



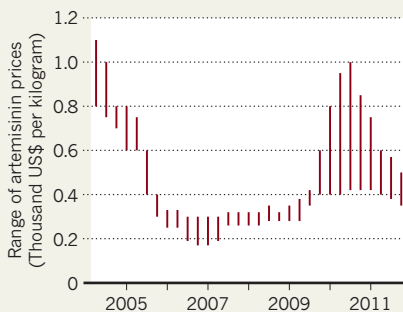
SOURCE: A2S2

But funding for ACTs is uncertain. In 2012, the AMFm subsidized about 60% of the world's total supply of the drugs. With an unusual approach — supplying most of the drugs through small, for-profit pharmacies, rather than public clinics — the programme got more ACTs into circulation. But donors were not convinced that the drugs were reaching the neediest patients. In November 2012, the Global Fund announced that it would merge the AMFm into its broader malaria campaign (see *Nature* **491**, 508; 2012). With so many demands on aid funding, some health experts worry that support for ACT distribution will falter.

Fabienne Jouberton, AMFm's senior officer for price negotiations, acknowledges that it will probably become much more difficult for *A. annua* growers and drug manufacturers to plan ahead. Rather than negotiating with a single body — the AMFm — they will have to assess demand separately in each individual country. "It's not optimal," she says.

### THE COST OF PROGRESS

Prices of the malaria drug artemisinin vary wildly. Introduction of an efficient, microbially produced version could make things worse.



For now, artemisinin costs about US\$400 per kilogram, and the semi-synthetic version is unlikely to be much cheaper. But Sanofi announced at the Nairobi meeting that it plans

to produce 60 tonnes of the compound in 2014, alarming *A. annua* growers and extractors, who said that the new supply could take more than one-third of their market. Sanofi and its partners are adamant that for the next several years, semi-synthetic artemisinin will be deployed only to smooth out market fluctuations, and will be sold at a 'no profit–no loss' price of \$350–400 that is unlikely to undercut other suppliers. The conference "almost felt like a price-fixing meeting," says Keasling wryly.

He is convinced that, eventually, the price could be lowered by making the fermentation process more efficient. Fitful funding is all the more reason to provide cheaper ACTs, adds Keasling, who would like all such drugs to be made using the semi-synthetic process. But he agrees that gradual introduction is necessary to avoid driving conventional producers out of business — at least "until we have enough installed capacity to take over the entire world supply". ■

### PUBLISHING

# Company offers portable peer review

*Author–pays service cuts down on redundant reviews.*

BY RICHARD VAN NOORDEN

Researchers waiting for their manuscript to emerge from multiple rounds of peer review as it bounces from journal to journal can easily get frustrated at the inefficiencies of the system. Soon they may have another option: paying for a fast, independent peer review that could travel with the paper from one journal to another.

The concept comes from a company called Rubriq. Charging authors an estimated US\$500–700 for its service, the firm plans to offer a standard-format anonymized review, and is currently testing its concept with publishers including Public Library of Science (PLOS), Karger, F1000Research and Wiley, as well as more than 500 reviewers. The company — owned by research software and services firm Research Square in Durham, North Carolina — is one of a number of firms hoping to capture functions usually performed by publishers.

Rubriq co-founder Keith Collier is confident that there is a gap in the market. "Millions of hours are spent every year on redundant reviews," he says. Collier has experience: he was previously general manager of ScholarOne, a widely used peer-review service based in Charlottesville, Virginia, and owned by Thomson

Reuters. He estimates that around 50% of articles are initially rejected. Every resubmission can lead to another round of peer review, delaying publication and costing time and money.

Recognizing the problem, journals are increasingly passing on their rejected papers with peer reviews attached; but such transfers tend to take place only within a publisher. Matthew Cockerill, managing director of BioMed Central (BMC) in London, says that BMC's *Genome Biology* accepts around 10% of submitted papers, but passes on 40% of those it rejects to other BMC journals together with the peer reviews; of those, around 50% are published somewhere within the BMC group.

Publishers are trying to share reviews more broadly: both BMC and PLoS are now working with the open-access journal *eLife* to explore passing on peer reviews with rejected papers, Cockerill says. But an early effort to do so — an alliance of neuroscience journals that in 2008 agreed to accept manuscript reviews from one another — has met with limited success. Before the scheme, journal editors who were wedded to their own peer-review standards "seemed more likely to use one another's toothbrushes than their review formats", says Clifford Saper, a former co-chair of the consortium. Even today, only 1–2% of rejected papers within the

group are shared with other journals, says John Maunsell, a current co-chair.

Collier is betting that journals will accept the standardized peer-review form it has created, which follows the example of a similar organization called Peerage of Science, based in Jyväskylä, Finland. That service has signed up a community of more than 1,100 scientists and has processed 67 manuscripts. Authors pay nothing. Instead, journals subscribe to the service and are charged up to €400 (US\$540) for each manuscript they accept (three have been published in this way so far). Scientists can also point other prospective publishers to the reviews of their papers.

Rubriq's emphasis is on speed. By paying peer reviewers \$100 each, it hopes to get reviews back within a week. That payment will not compensate a reviewer for his or her time (which economic analyses have estimated as worth on the order of \$400 per review) but it could start to professionalize what has historically been a voluntary role, Collier says. The payments are covered by Rubriq's fee to authors, which also covers the administrative burden of recruiting reviewers and assigning papers — an estimated \$200. Pete Binfield, publisher of *PeerJ*, which aims to cut the costs of publishing dramatically and which launched its first articles this week, says that \$200 is a fair cost, and thinks that those activities cost about the same at *PeerJ*. (Binfield is on Rubriq's advisory panel.)

Other publishers are taking a more radical approach to publishing models: from the European Molecular Biology Organization's open peer review, which is not anonymous, to F1000Research's strategy: publish first and peer review later. But Collier says that Rubriq aims "to be an independent validation service to streamline the publishing process; we're not trying to disrupt the industry". ■





# NEW YORK VS THE SEA

BY JEFF TOLLEFSON

*In the wake of Hurricane Sandy, scientists and officials are trying to protect the largest US city from future floods.*

Joe Leader's heart sank as he descended into the South Ferry subway station at the southern tip of Manhattan in New York. It was 8 p.m. on 29 October, and Hurricane Sandy had just made landfall some 150 kilometres south in New Jersey. As chief maintenance officer for the New York city subway system, Leader was out on patrol. He had hoped that the South Ferry station would be a refuge from the storm. Instead, he was greeted by wailing smoke alarms and the roar of gushing water. Three-quarters of the way down the final set of stairs, he pointed his flashlight into the darkness: seawater had already submerged the train platform and was rising a step every minute or two.

"Up until that moment," Leader recalls, standing on the very same steps, "I thought we were going to be fine."



Opened in 2009 at a cost of US\$545 million, the South Ferry station is now a mess of peeling paint, broken escalators and corroded electrical equipment. Much of Manhattan has returned to normal, but this station, just blocks from one of the world's main financial hubs, could be out of service for 2–3 years. It is just one remnant of a coastal catastrophe wrought by the largest storm in New York's recorded history.

Sandy represents the most significant test yet of the city's claim to be an international leader on the climate front. Working with scientists over the past decade, New York has sought to gird itself against extreme weather and swelling seas and to curb emissions of greenhouse gases — a long-term planning process that few other cities have attempted. But Sandy laid bare the city's vulnerabilities, killing 43 people, leaving thousands homeless, causing an estimated \$19 billion in public and private losses and paralysing the financial district. The New York Stock Exchange closed for the first time since 1888, when it was shut down by a massive blizzard.

As the humbled city begins to rebuild, scientists and engineers are trying to assess what happened during Sandy and what problems New York is likely to face in a warmer future. But in a dilemma that echoes wider debates about climate change, there is no consensus about the magnitude of the potential threats — and no agreement about how much the city should spend on coastal defences to reduce them.

On 6 December, during his first major public address after the storm, New York mayor Michael Bloomberg promised to reinvest wisely and to pursue long-term sustainability. But he warned: “We have to live in the real world and make tough decisions based on the costs and benefits.” And he noted that climate change poses threats not just from flooding but also from drought and heat waves. The city must be mindful, he said, “not to fight the last war and miss the new one ahead”.

## CALCULATED RISKS

In the immediate aftermath of Sandy, lower Manhattan looked like a war zone. Each night, streams of refugees wielding flashlights wandered north out of the blackout zone, where flood waters had knocked out an electrical substation.

The storm devastated several other parts of the city as well. In Staten Island, pounding waves destroyed hundreds of homes, and one neighbourhood in Queens burned to ashes after water sparked an electrical fire. Power outages lasted for more than two weeks in parts of the city. Chastened by the flooding and acutely aware that Hurricane Irene, in 2011, was a near miss, the city is now wondering what comes next.

“Is there a new normal?” asks John Gilbert, chief operating officer of Rudin Management, which

➔ **NATURE.COM**  
For maps and a  
slideshow about  
Sandy, see:  
[go.nature.com/Sejine](http://go.nature.com/Sejine)

manages several office buildings in downtown New York. “And if so, what is it?” Gilbert says that the company is already taking action. At one of its buildings, which took on some 19 million litres of water, the company is moving electrical systems to the second floor. “You have to think that as it has happened, it could happen again,” he says. “And it could be worse.”

At Battery Park, near the South Ferry station, the storm surge from Sandy rose 2.75 metres above the mean high-water level — the highest since gauges were installed there in 1923. In a study published last week in *Risk Analysis*, researchers working with data from simulated storms concluded that a surge of that magnitude would be expected to hit Battery Park about once every 500 years in the current climate (J. C. J. H. Aerts *et al.* *Risk Anal.* <http://dx.doi.org/10.1111/risa.12008>; 2013).

**“IT COULD HAPPEN AGAIN, AND IT COULD BE WORSE.”**

org/10.1111/risa.12008; 2013).

But the study authors and other scientists say that the real risks may be higher. The study used flooding at Battery Park as a measure of hurricane severity, yet it also showed that some storms could cause less damage there and still hammer the city elsewhere. Factoring in those storms could drive up the probability estimates of major hurricane damage to New York.

The 1-in-500 estimate also does not take into account the unusual nature of Sandy. Dubbed a Frankenstorm, Sandy was a marriage of a tropical cyclone and a powerful winter snowstorm, and it veered into the New Jersey coast along with the high tide of a full Moon. “It was a hybrid storm,” says Kerry Emanuel, a hurricane researcher at the Massachusetts Institute of Technology (MIT) in Cambridge and one of the study's co-authors. “We need to understand how to assess the risks from hybrid events, and I'm not convinced that we do.”

The risks will only increase as the world warms. The New York City Panel on Climate Change's 2010 assessment suggests that local sea level could rise by 0.3–1.4 metres by 2080. Last year, Emanuel and his colleagues found that floods that occur once every 100 years in the current climate could happen every 3–20 years by the end of this century if sea level rises by 1 metre. What is classified as a ‘500-year’ event today could come every 25–240 years (N. Lin *et al.* *Nature Clim. Change* 2, 462–467; 2012).

For city planners, the challenge is to rebuild and protect the city in the face of scientific uncertainty. A few scientists have said for more than a decade that the city should armour New York's harbour with a storm-surge

barrier similar to the Thames barrier in London. In Sandy's wake, that idea has gained renewed interest, and a New York state panel last month called for a formal assessment of it.

## BRIDGES AND BARRIERS

Malcolm Bowman, who heads the storm-surge modelling laboratory at the State University of New York at Stony Brook, has spearheaded the drive for barriers. He imagines a structure roughly 8 kilometres wide and 6 metres high at the entrance to the harbour, and a second barrier where the East River drains into the Long Island Sound. The state panel's cost estimates for such a system range from \$7 billion to \$29 billion, depending on the design. The harbour barrier could also serve as a bridge for trains and vehicles to the city's airports, suggests Bowman. “My viewpoint is not that we should start pouring concrete next week, but I do think we need to do the studies,” he says. But whether Sandy will push the city to build major defences, Bowman says, “I don't know.”

Disasters have spurred costly action in the past. The 1888 blizzard helped to drive New York to put its elevated commuter trains underground. And in 2012, the US Army Corps of Engineers completed a \$1.1-billion surge barrier in New Orleans, Louisiana, as part of a \$14.6-billion effort to protect the city after it was battered by hurricanes Katrina and Rita in 2005. But the New York metropolitan area is bigger and more complex than New Orleans, and protecting it will require a multi-pronged approach. Several hundred thousand city residents live along more than 800 kilometres of coastline, and a barrier would not protect much of coastal Long Island, where Sandy wrought considerable damage. Moreover, the barrier would work only against occasional storm surges. It would not hold back the slowly rising sea or protect against flooding caused by rain.

“A storm-surge barrier may be appropriate, but it's never one thing that is going to protect you,” says Adam Freed, a programme director at the Nature Conservancy in New York, who until late last year was deputy director of the city's office of long-term planning and sustainability. “It's going to be a holistic approach, including a lot of unsexy things like elevating electrical equipment out of the basement and providing more back-up generators.”

As part of that holistic effort, officials are exploring options for expanding the remaining bits of wetlands that once surrounded the city and buffered it from storms. In his address, Bloomberg called wetlands “perhaps the best natural barriers against storms that we have”.

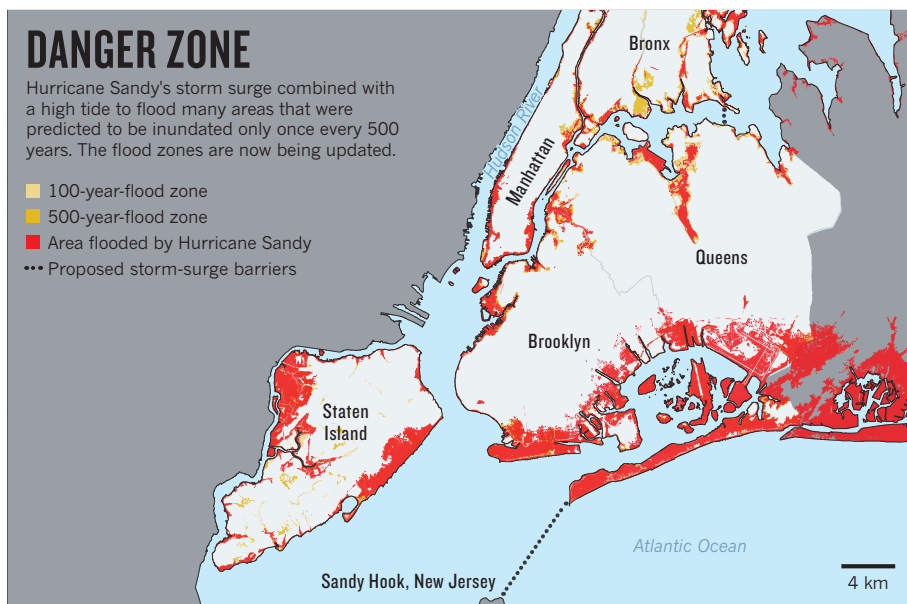
But most of the city's wetlands have become prime real estate in recent decades, and Sandy made clear the consequences of developing those areas, says Marit Larson, director of wetlands and riparian restoration for the New York parks department.

A few weeks after the storm, Larson parks her

## DANGER ZONE

Hurricane Sandy's storm surge combined with a high tide to flood many areas that were predicted to be inundated only once every 500 years. The flood zones are now being updated.

- 100-year-flood zone
- 500-year-flood zone
- Area flooded by Hurricane Sandy
- Proposed storm-surge barriers



car near the beach on Staten Island and looks out at a field of *Phragmites australis*, a common marsh reed. The field is part of Staten Island's 'Bluebelt' programme, initiated in the late 1980s to promote wetlands and better manage storm-water runoff. But the patch of wetlands here is smaller than a football pitch, and Sandy's surge rolled over it, damaging the nearby row houses. "If you look at the historical maps," says Larson, "everything that used to be a wetland got wet."

### UP TO CODE

New York is now moving to strengthen its network of existing wetlands, which cover some 2,300–4,000 hectares. The mayor's budget plan for 2013–17 includes more than \$200 million to restore wetlands as part of an effort to protect and redesign coastal developments.

Sandy also showed how proper construction can help to reduce risks from future storms. In one Staten Island neighbourhood, a battered roof rests on the ground, marking the spot where an ageing bungalow once stood. Next door, a newer house still stands, with no apparent damage apart from a flooded garage — sturdy proof of the value of modern building codes. In New York, newer buildings constructed in 100-year-flood zones, which are defined by the US Federal Emergency Management Agency (FEMA), cannot have any living spaces or major equipment, such as heating units, below the projected flood level (see 'Danger zone').

The city's zoning provisions could not protect against a storm like Sandy: officials estimate that two-thirds of the homes damaged by the storm were outside the 100-year-flood area. But scientists say that the FEMA flood maps were out of date, so even century-scale storms could cause damage well beyond the designated areas. Last month, FEMA began releasing new flood maps for the New York region that substantially expand this zone.

In their latest study, Emanuel and his colleagues estimate the average annual flood risk for New York as only \$59 million to \$129 million in direct damages. But costs could reach \$5 billion for 100-year storms and \$11 billion for 500-year storms. These figures do not include lost productivity or damage to major infrastructure, such as subways.

Bowman and other researchers argue that the city should commit to protecting all areas to a 500-year-flood standard, but not all the solutions are physical. A growing chorus of academics and government officials stress that the city must also bolster its response capacity and shore up the basic social services that help people to rebuild and recover.

**"ULTIMATELY, WE ALL HAVE TO MOVE TOGETHER TO HIGHER GROUND."**

Most importantly, the city and surrounding region need to develop a comprehensive strategy for defending the coastline, says Jeroen Aerts, a co-author of the *Risk Analysis* assessment who studies coastal-risk management at VU University in Amsterdam. Aerts is working with New York officials to analyse proposals for the barrier system and a suite of changes in urban planning, zoning and insurance. "You need a master plan," he says.

Seth Pinsky is working towards that goal. As president of the New York City Economic Development Corporation, he was tapped by Bloomberg to develop a comprehensive recovery plan that will make neighbourhoods and

infrastructure safer. He points out that some newer waterfront parks and residential developments along the coast fared well during the storm. For example, at Arverne by the Sea, a housing complex in Queens, Pinsky says that units survived because they are elevated and set back from the water, with some protection from dunes. The buildings suffered little damage compared with surrounding areas.

### INTELLIGENT DESIGN

The cost of strengthening the city will be astronomical. In January, Congress approved some \$60 billion to fund Sandy recovery efforts, with around \$33 billion for longer-term investments, including infrastructure repair and construction by the Army Corps of Engineers. Pinsky says that he does not yet know how much of that money will go to New York, but he is sure it will not be enough. The city will define its budget in June, after his group has made its official recommendations. The rebuilding endeavour will probably necessitate a "creative" mix of public and private financing, he says. "It will probably require calling on a combination of almost every tactic that has been tried around the world."

Even as he calls for more intelligent development, Pinsky says that New York is unlikely to take a drastic approach to dealing with storm surge and sea-level rise. "Retreating from the coastline of New York city both will not be necessary and is not really possible," he says.

Given the sheer scale of development along the coast, it is hard to argue with Pinsky's assessment. But many climate scientists fear that bolstering coastal developments only delays the eventual reckoning and increases the likelihood of future disasters. The oceans will rise well into the future, they say, so cities will eventually be forced to accommodate the water.


"I don't see anything yet that looks towards long-term solutions," says Klaus Jacob, a geoscientist at Columbia University's Lamont-Doherty Earth Observatory in Palisades, New York. But Jacob admits that he is as guilty as anyone. In 2003, he and his wife bought a home in a low-lying area on the Hudson River in Piermont, New York. Although it went against his professional principles, he agreed to the purchase with the assumption that he could elevate the house. But height-restriction laws prevented him from doing so, and Sandy flooded the house. The couple are now rebuilding.

"In a way, I think I was in denial about the risk," Jacob says. He hopes that a new application to raise the house will be approved, but he still fears that the neighbourhood will not survive sea-level rise at the end of the century. New Yorkers and coastal residents everywhere would be wise to learn that lesson. "Ultimately," Jacob says, "we all have to move together to higher ground." ■ [SEE EDITORIAL P.148](#)

**Jeff Tollefson** reports for Nature from New York.

SOURCES: FEMA; CITY OF NEW YORK; CUNY





By offering unproven therapies, a Texas biotechnology firm has sparked a bitter debate about how stem cells should be regulated.

# COWBOY CULTURE

BY DAVID CYRANOSKI

**A**nn McFarlane is losing faith. In the first half of 2012, the Houston resident received four infusions of adult stem cells grown from her own fat. McFarlane has multiple sclerosis (MS), and had heard that others with the inflammatory disease had experienced improvements in mobility and balance after treatment. The infusions — which have cost her about US\$32,000 so far — didn't help, but she knew that there were no guarantees.

It is McFarlane's experience with Celltex Therapeutics, the company that administered the cells, that bothers her. She was told that she had been enrolled in a study to test the cells' efficacy, but received almost no information about it. And although it wasn't exactly a secret that the treatment had not been approved by the US Food and Drug Administration (FDA), Celltex, based in Houston, Texas, assured its clients that it was within its rights to provide it. But Celltex was forced to halt treatments in October, and in November a legal battle broke out over who owned the cells still being stored by the

company. For weeks, McFarlane was uncertain whether her cells were being grown and stored properly. Although Celltex has told its customers that it has settled the dispute, McFarlane has her doubts. "I am not confident that the cells are viable and safe," she says. "I probably will not feel comfortable using these cells."

For the past decade, people such as McFarlane have searched far and wide for clinics offering to deliver on the promise of adult stem cells. Unlike embryonic stem cells, their use does not require the controversial destruction of an embryo. Yet although adult stem cells are claimed to ameliorate a wide range of disorders, they have not yet been shown to do so conclusively in clinical trials in the United States. Relying on customer testimonies and company promises, patients have travelled to clinics in places such as China, Costa Rica, Mexico and Japan to receive them from unregulated, often unaccredited, laboratories, driving a boom in stem-cell tourism. According to Leigh Turner, a bioethicist at

the University of Minnesota in Minneapolis, at least ten clinics offer treatments in the United States. Turner and others have questioned the quality of the cells that these firms provide, and several outlets have been forced to stop providing treatments.

CellTex has been one of the most visible. Established in 2011 (see 'Texas throwdown'), it offered therapies for conditions as varied as arthritis, back pain, MS and Parkinson's disease. It produced its cells in a 1,400-square-metre, state-of-the-art facility in Sugar Land, Texas, that was registered with the FDA, and — the company claimed — had strict quality control. Although the company flouted federal regulations, which deem such cell therapies to be biological drugs, it adhered to state rules, which had recently been tailored to bolster the stem-cell industry. But even with the support of Texas governor Rick Perry — Celltex's first patient — the company had to stop offering treatments in the United States.

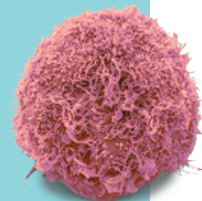
Throughout the ordeal, Celltex has stirred up a polarizing public debate. Channelling the state's ethos of rugged individualism, Stanley

OLIVER MUNDAY

**NATURE.COM**  
To hear more about  
this story, visit:  
[go.nature.com/ac3ufp](http://go.nature.com/ac3ufp)

# TEXAS THROWDOWN

**2006** Seoul-based RNL Bio devises stem-cell extraction and cultivation methods. Delivery of cells to patients, prohibited in South Korea, is done in Japan and China.



**May 2010** Orthopaedic surgeon Stanley Jones travels to Japan for stem cells to treat his arthritis.

**March 2011** Jones and executive David Eller ink US\$30-million deal to license RNL Bio's methods in the United States and create Celltex in Texas.

**July 2011** Texas governor Rick Perry (pictured), who had backed legislation to support stem-cell treatments in Texas, becomes Celltex's first patient.



**February 2012** Bioethicist Glenn McGee resigns as adviser to Celltex amid controversy.

**April 2012** The US Food and Drug Administration (FDA) finds problems at Celltex's cell-processing plant. The Texas Medical Board sets rules allowing Celltex to bypass the FDA.

**July 2012** Six California patients who received stem-cell treatments from RNL Bio sue the company for misrepresenting the benefits.

**September 2012** The FDA warns Celltex that its treatments require federal approval before use in patients.

**October 2012** Celltex tells patients that it must suspend stem-cell treatments.

**November 2012** Celltex and RNL Bio sue each other over payments for storing and maintaining cells.

**January 2013** Celltex announces that it will restart its treatments, in Mexico, and will meet with the FDA in March.

Jones, the orthopaedic and cosmetic surgeon who founded Celltex, argued that the federal government was overstepping its bounds. The company's therapies, he said, are not biological drugs, but tissue transplants, which typically do not require clinical trials to prove that they work. The drama raised and dashed the hopes of patients, many of whom rallied to support Celltex. Now, some in the industry worry that the turmoil will tar the entire field. "It doesn't serve the interests of patients or the medical community at large," says Gil Van Bokkelen, head of Athersys, a biotech firm in Cleveland, Ohio, that has been cooperating with the FDA to test stem-cell treatments in clinical trials.

Celltex has declined to comment on its business decisions, but it has been promising its patients a reboot, saying that it will publish the results of preliminary studies; start the FDA drug approval process so that it can resume treatments in the United States; and make its stem-cell treatments available in Mexico (see *Nature* <http://doi.org/kdd>; 2013). Although many former patients continue to support the company's endeavours, some, such as McFarlane, see the firm as an unfortunate distraction. "Any bad stem-cell PR is detrimental to the overall picture, which is positive," she says.

## HOME-GROWN THERAPY

In an earnest Southern drawl, Jones recounted how he co-founded Celltex. He was one of the headline speakers at the first Houston Stem Cell Summit at the Houstonian Hotel, a plush resort surrounded by 7 hectares of woods. Jones said that he experienced the miracle of stem cells first-hand in 2010 when he travelled to Japan to receive infusions to treat his crippling arthritis. His cells came from RNL Bio, a stem-cell company based in Seoul. Within three months, he said, he could do pilates and work as hard as he had in his youth.

But it irked Jones that he and others had to board a plane to get better. So, with the help of David Eller, a former executive of DuPont in Europe and a friend of governor Perry, he gathered the \$30 million he needed to gain exclusive rights from RNL to use its stem-cell processing and banking technology in the United States.

With no background in stem cells, "I had to start from scratch," Jones recalled. "I knew that I was the only game in town. So I learned to do liposuction, and I started doing liposuction on anybody that wanted stem cells." Perry, a charismatic politician and one-time presidential hopeful, became Jones' first patient in July 2011, seeking treatment for his recurring back pain. Jones said that he tried to maintain a low profile for the company — not even establishing a website for months, relying instead on personal recommendations — but attention soon followed.

Operating in a state that has historically bristled under big government, Celltex appealed not only to desperate patients but also to the ideals of the political and religious right and their opposition to big government and embryonic

stem-cell research. It campaigned for its cause at intimate 'educational forums' at the Houstonian, arguing that it could already provide ethically acceptable treatments with adult stem cells. Patients spread the word at churches, and bloggers praised the company for asserting patients' rights. When prospective clients showed up in his office, Jones would tell them of his experience; at times, he reportedly wept. By October 2012, 233 people had been treated with Celltex stem cells.

Celltex offered treatments with 'mesenchymal stem cells' (MSCs), which are found in bone marrow, muscle, fat and other tissue. In the procedure licensed from RNL, the cells are extracted from just five grams or so of fatty tissue, then cultured, expanded and banked. Doctors could then administer the cultured stem-cell concoction intravenously.

RNL developed its method for isolating and culturing MSCs in 2006, and within four years, 7,000 patients had received its cells — for disorders including stroke, renal failure, asthma, psoriasis and MS. The recoveries — recounted in anecdotes on its website and in the book *The Grace of Stem Cells* by RNL chairman Jeong-Chan Ra (Tate Publishing, 2012) — are often complete and miraculous. Some patients have the treatment for fatigue, to remove wrinkles or for "restoring the skin's elasticity", Ra writes in his book. But even in South Korea, which seems to have approved more adult-stem-cell treatments than any other country, none of RNL's treatments has been approved. That is why Jones and others had to travel to Japan or China to receive them.

MSCs do have therapeutic potential. Research over the past ten years has suggested that they home in on damaged or inflamed tissues and have various roles in repair, releasing molecules that suppress an overactive immune system, stabilize newly forming blood vessels or prevent cells from dying<sup>1</sup>. Initial fears that they might take hold and become cancerous subsided with evidence that they disappear after a short time in the body<sup>2</sup>. There have been reports of complications, such as "catastrophic demyelinating encephalomyelitis", a life-threatening inflammation of the brain<sup>3</sup>. But there is a growing consensus that MSCs are safe<sup>4</sup>.

Arnold Caplan from Case Western Reserve University in Cleveland, Ohio, did pioneering work on MSCs and refers to their action as "hit and run" healing. More than 250 MSC trials are ongoing or completed, he told the audience at the October summit. And even though most of these studies are small and none has led to regulatory approval for widespread use in the United States or Europe, Caplan says that he would use the cells. "If I had MS, I would be getting this therapy. I'd probably go offshore." That is not a view endorsed by the International Society for Stem Cell Research in Skokie, Illinois, however, which urges prospective patients to be cautious about stem-cell tourism.

Safety also depends on how cells are handled,

STEVE GSCHNEISSER/SPL

LUCIAN PERKINS FOR THE WASHINGTON POST/GETTY



and the FDA has voiced concerns about Celltex's operations. In April 2012, an FDA inspection found a laundry list of problems — 79 in all — at the Sugar Land plant that the firm ran with RNL. The list included failure to confirm that equipment was sterile, and operating table-top centrifuges and incubators on the floor. Labels on one bottle said different things in Korean and English and one product used in the culture media was marked “not intended for human or animal diagnostic or therapeutic uses”.

Celltex blamed a “language barrier”. In a formal statement, it said it was maintaining an “open line of communication” and a “cooperative relationship” with the FDA and that it would address the problems. “Celltex and its partner RNL Bio process stem cells in a safe, sterile laboratory with procedures that ensure cell viability and integrity,” the company said.

But the FDA also found fault with the therapy itself — in particular, it challenged Jones's claim that the procedure qualifies as a treatment with a patient's own cells and therefore falls outside FDA jurisdiction. To qualify as such, the cells would have to be “minimally manipulated” and be implanted for “homologous use”, meaning that they carry out the same functions in the treated tissue as they do in the tissue from which they are extracted.

On 24 September, the FDA sent a warning letter to Celltex saying that because the firm's processing “alters the original relevant characteristics” of the fat tissue, the cells are not considered minimally manipulated. And because the fat-derived MSCs were used to treat problems in other tissue, such as nerve tissue in MS, it does not count as homologous use. The FDA demanded that Celltex comply and show that it had fixed all the manufacturing problems and that it had approval to use the stem cells, or face regulatory action, which could include seizure of the facility or an injunction.

Celltex responded to the FDA on 16 October, again arguing that the preparation “does not alter the relevant biological characteristics of the MSCs”. It also claimed that MSCs are known to spur the growth of blood vessels, regulate the immune response and dampen inflammation, and that those are presumably the roles the cells carry out in both the transplant and the original fat tissue.

Peter Connick of the University of Cambridge, UK, who is running clinical studies with MSCs, says that he suspects that the FDA is correct that MSCs function differently in different tissue. “The ability of MSCs to instruct neural stem-cell fate decisions, for example, is by definition not something that they could do in their bone-marrow niche,” he says.

On 25 October, Celltex notified its customers that it had to halt treatments while it prepared to start an FDA-approved clinical trial. The message was apologetic, but cast the blame on the FDA. “I know that many of you are deeply hurting due to the FDA's decision to regulate your own stem cells as a drug,” it read.

That sentiment was reflected at the October summit, which occurred ten days later. Sponsored by Celltex, the conference had attracted representatives from many of the country's most promising stem-cell companies and banks — including those that have been working through the FDA approval process. Perry spoke, praising the “wildcat spirit” in Texas and bemoaning the FDA's “crippling bureaucratic infrastructure”. Lawyers and ethicists debated the need for regulations on stem cells along-

## “Any bad stem-cell PR is detrimental to the overall picture.”

side patients, some of whom were wheelchair-bound. “The FDA is trying to screw you,” one participant told Eller during a coffee break.

Yet many in the industry back the FDA's position. Van Bokkelen of Athersys was the previous head of the Alliance for Regenerative Medicine in Washington DC, which represents the interests of 100 or so US companies, many of which, like Athersys, have stem-cell treatments undergoing FDA-approved clinical trials. He says that stem-cell treatments may require changes in existing regulations, but that he hopes to work these out collaboratively with the FDA. “Rather than coming at them with pitchforks, I wanted to engage in dialogue to come up with ways to explore areas of uncertainty,” he says.

### FRUSTRATED PATIENTS

Mary Pat Moyer, the founder of INCELL, a life-science services company based in San Antonio, Texas, started her presentation at the October summit with a slide that criticized the “stem-cell cowboy” mentality that Celltex has espoused. Frustrated Celltex customers had contacted her about storing their cells in her repository (she refused), and she says that she was surprised to hear how little data the company had collected on patients from ongoing tests. “It's frustrating. If they have so many patients, they should provide some guidance” in terms of what dosing works best for each disease.

Despite being part of a study, McFarlane notes that Celltex didn't ask her about her progress on her treatment until January this year. She had hounded the company for more information about her treatments. Celltex showed, she says, a “total lack of follow-up or investigation”.

Many researchers are pursuing more rigorous trials. For example, Connick is leading a clinical trial of stem cells to treat MS in which he is using various neurological tests as well as techniques to show improvement in the neurons. Connick says that MSC therapy seems to be “feasible, safe and may slow the neuroaxonal loss” that leads to the progressive disability in MS<sup>5</sup>. He does concede, however, that “questions remain about the

long-term safety and long-term efficacy”, and that he would not advise patients to try the treatment outside of clinical trials.

In November 2012, Celltex's troubles deepened. RNL Bio and Celltex sued each other over how much money Celltex needed to pay for the services that RNL was providing at the Sugar Land plant. Celltex had to file a restraining order to get access to the cells.

But in the months since the company halted treatments, Celltex has said that it is making overtures to the FDA, and that it will now collect information from its clinical studies. In January, it told patients that a clinical research organization was evaluating data and that patients should expect a follow-up call about their conditions. It also told its patients that it would start approved clinical trials “shortly after” a meeting with the FDA in March. And it announced that it would be providing treatments through physicians in Mexico.

Most patients, meanwhile, are frustrated by the delays. Debbie Bertrand, who has MS, said that, before going to Celltex, she was turned away from five approved clinical trials because her symptoms weren't severe enough to qualify. “I don't want to wait till I'm bad enough,” she says. And many are mistrustful of the FDA, suspecting that it is in cahoots with big drug companies to peddle ineffective treatments. “I have tried all of the FDA-approved MS drugs, which had horrible life-threatening side effects,” says Sammy Jo Wilkinson, an MS patient and Celltex client who gave a moving account of her ordeal at the October summit.

Other companies now seem poised to fill the void left by Celltex, albeit less brazenly. According to Turner, who has followed closely the proliferation of companies providing such treatments in the United States, “businesses are becoming more cautious about the language they use on their websites. Many of them now emphasize that they engage in no more than minimal manipulation of stem cells before administering them.” He adds, “In my opinion, the reason the FDA's letter to Celltex does not seem to have had a major deterrent effect on other companies is that marketing and administering autologous stem cells is so profitable.”

McFarlane, for one, still hopes to find a stem-cell therapy that works for her. “Even though I'm disenchanted with Celltex, I still hold out great hope for adult stem-cell treatment in general. I personally have not been helped ... but I have seen some amazing results for others.” ■ [SEE EDITORIAL P.147](#)

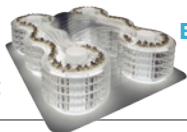
David Cyranoski writes for *Nature* from Tokyo.

1. Caplan, A. I. & Correa, D. *Cell Stem Cell* **9**, 11–15 (2011).
2. von Bahr, L. et al. *Stem Cells* **30**, 1575–1578 (2012).
3. Alderazi, Y. J. et al. *J. Child Neurol.* **27**, 632–635 (2012).
4. Lalu, M. M. et al. *PLoS ONE* **7**, e47559 (2012).
5. Connick, P. et al. *Lancet Neurol.* **11**, 150–156 (2012).



# COMMENT

**ECODESIGN** Olympic velodrome engineer builds with nature **p.172**



**ECODESIGN** Materials makers on how to do more with less **p.174**

**THEATRE** New York play explores why Isaac Newton stuck a needle in his eye **p.175**

**METRICS** Some altmetrics are too easy to game so lack credibility **p.176**

DIMITAR DILKOFF/AFP/GETTY



Volunteer cleaners negotiate a Bulgarian reservoir jammed with plastics.

## Classify plastic waste as hazardous

Policies for managing plastic debris are outdated and threaten the health of people and wildlife, say **Chelsea M. Rochman, Mark Anthony Browne** and colleagues.

Last year, 280 million tonnes of plastic was produced globally. Less than half of it was consigned to landfill or recycled. Of the remaining 150 million tonnes, some may still be in use; the rest litters continents and oceans (see 'Plastic world').

Plastic debris can physically harm wildlife<sup>1,2</sup>. Moreover, many plastics may be chemically harmful in some contexts — either because they are themselves potentially toxic<sup>3</sup> or because they absorb other pollutants<sup>4,5</sup>. Yet in the United States, Europe,

Australia and Japan, plastics are classified as solid waste — so are treated in the same way as food scraps or grass clippings.

We believe that if countries classified the most harmful plastics as hazardous, their environmental agencies would have the power to restore affected habitats and prevent more dangerous debris from accumulating. Ultimately, such a move could boost research on new polymers and replace the most problematic materials with safer ones.

It is now almost impossible to walk in

the countryside or on a beach without encountering bits of plastic. Larger pieces, from bottles and bags to floating pontoons, can transport species to new habitats where they might do damage. Such debris can kill or injure ecologically and commercially important species, including mussels, salt-marsh grasses and corals<sup>1,2</sup>. Mammals, reptiles and birds can also be harmed through eating plastic or becoming entangled in it. Last year, the secretariat of the Convention on Biological Diversity in Montreal, Canada, reported ►



► that all sea turtle species, 45% of marine mammal species and 21% of seabird species can be harmed in this way.

### HEALTH RISK

As plastic breaks into smaller pieces, it is more likely to infiltrate food webs<sup>2</sup>. In laboratory and field studies, fish, invertebrates and microorganisms ingest micrometer-sized particles<sup>2</sup>, which also come from synthetic (polyester or acrylic) clothing<sup>6</sup> and cleaning products containing plastics. More research is needed to investigate the effects of organisms ingesting debris in the wild. Nevertheless, studies in humans<sup>7</sup> and mussels<sup>2</sup> have found that ingested and inhaled microplastic gets into cells and tissues where it can cause harm. (In patients who have had their knee or hip joints replaced with plastic implants, such particles can disrupt cellular processes and degrade tissues.)

Plastics are made up of repeating units called monomers that bind together to form long chains, or polymers. These chains are generally thought to be chemically inert, yet unreacted monomers and other harmful ingredients can be found in plastics<sup>3,4</sup>. According to a hazard-ranking model based on the United Nations' Globally Harmonized System of Classification and Labeling of Chemicals, the chemical ingredients of more than 50% of plastics are hazardous<sup>3</sup>. Studies investigating, for instance, the transfer of additives in polyvinylchloride (PVC) from medical supplies to humans indicate that these chemicals can accumulate in the blood<sup>8</sup>. In laboratory tests, monomers and other ingredients of PVC, polystyrene, polyurethane and polycarbonate can be carcinogenic and can affect organisms in a similar way to the hormone oestrogen<sup>3,4,9</sup>.

The monomers making up some plastics, such as polyethylene (used to make carrier bags), are thought to be more benign. Yet these materials can still become toxic by picking up other pollutants<sup>4,5</sup>. Pesticides and organic pollutants such as polychlorinated biphenyls are consistently found on plastic waste at harmful concentrations 100 times those found in sediments and 1 million times those occurring in sea water<sup>4</sup>. Many of these are 'priority pollutants': chemicals that are regulated by government agencies, including the US Environmental Protection Agency (EPA), because of their toxicity or persistence in organisms and food webs. These chemicals can disrupt key physiological processes, such as cell division and immunity, causing disease or reducing organisms' ability to escape from predators or reproduce.

In an analysis (unpublished results), we

found that at least 78% of priority pollutants listed by the EPA and 61% listed by the European Union are associated with plastic debris. Some are ingredients of plastic, and others are absorbed from the environment. Preliminary evidence indicates that priority pollutants can enter the tissues of species after they eat debris<sup>4,10</sup>. Seabirds that have consumed plastic waste have polychlorinated biphenyls in their tissues at 300% greater concentrations than in those that have not eaten plastic<sup>4</sup>.

### NAME GAME

Governments have struggled for decades to reduce plastic debris. The International Convention for the Prevention of Pollution From Ships (MARPOL) was signed in 1973, although a complete ban on the disposal of plastics at sea was not enacted until the end of 1988. Yet despite 134 nations agreeing to eliminate plastics disposal at sea, oceanic sampling suggests that the problem has persisted or worsened since MARPOL was signed. In the North Pacific, the concentration of microplastic debris has increased by two orders of magnitude. As far as we know, no attempts have been made to regulate the disposal of plastics on land at an international level.

We feel that the physical dangers of plastic debris are well enough established, and the suggestions of the chemical dangers sufficiently worrying, that the biggest producers of plastic waste — the United States, Europe and China — must act now. These countries should agree to classify as hazardous the most harmful plastics, including those that cannot be reused or recycled because they lack durability or contain mixtures of materials that cannot be separated.

Focusing on the most problematic materials is a realistic first step. Currently, just four plastics — PVC, polystyrene, polyurethane and polycarbonate<sup>3,4</sup> — make up roughly 30% of production. These are

particularly difficult to recycle and are made of potentially toxic materials. PVC is used in construction, such as in pipes that carry drinking water; polystyrene is used for food packaging; polyurethane is used for furniture; and polycarbonate in electronics. Healthcare and technology industries are already replacing PVC components in intravenous-drip bags and in computers with materials that are safer, more durable and recyclable, such as polypropylene and aluminium.

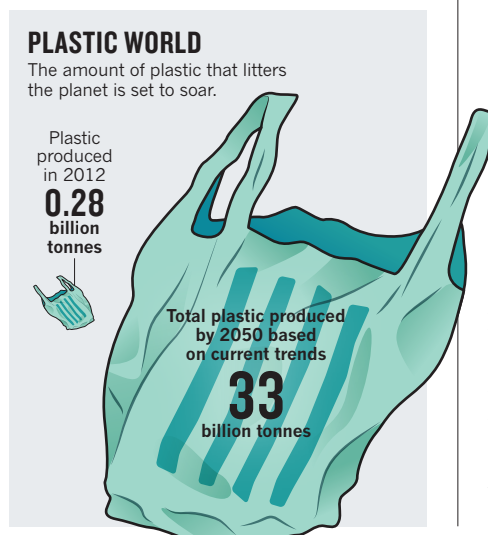
With a change in plastics categorization, numerous affected habitats could immediately be cleaned up under national legislation using government funds. In the United States, for instance, the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 would enable the EPA to clear the vast accumulations of plastic that litter the terrestrial, freshwater and marine habitats under US jurisdiction.

### CHAIN REACTION

History shows that this approach works. Chlorofluorocarbons (CFCs) and persistent organic pollutants were reclassified as hazardous under the Montreal Protocol in 1989 and the Stockholm Convention in 2004, respectively. This led, in each case, to nearly 200 countries stopping the production of some 30 dangerous chemical groups and replacing them with safer ones. For CFCs, all production stopped within seven years.

Our critics counter that without evidence of catastrophic harm to health or the environment, it is a stretch to equate plastics to CFCs and other substances classed as toxic. We disagree. We believe that manufacturers of plastic, along with the food and textile industries that rely heavily on it, should have to prove that their products and packaging are safe. Such demands are routinely made on the food and pharmaceutical industries by directives from numerous agencies, including the US Food and Drug Administration and the European Medicines Agency.

Ultimately, changes in regulation need to drive the development of a closed-loop system in which all plastics are reused and recycled. Today, most plastic waste goes to landfills where chemicals leach from the plastic into surrounding habitats<sup>4</sup>. Worldwide, the recycling of plastics is increasing. From 2005 to 2010, for example, US and UK plastics recycling rose by 4% and 9%, respectively. Still, current efforts to 'reduce, reuse, recycle' cause other problems. Recycling often involves burning plastics and using the energy released for other purposes, but incineration can generate priority pollutants and greenhouse gases. In a closed-loop system, plastics would be continually reused and replenished only when materials become too degraded — analogous to the reuse of glass bottles by the





Wildlife such as this white stork (*Ciconia ciconia*) can become entangled in discarded plastic bags.

UK dairy industry from the late 1800s to the mid-1990s.

Many people think that replacing materials such as wood and glass with plastic to make goods lighter can help to address climate change. However, the benefits must be balanced against the negative impacts of plastics so that they are used only when they have smaller carbon and ecological footprints than alternatives. Others may argue that in the current global economic crisis, nations can ill afford to regulate an industry that, in the United States alone, is worth US\$1 trillion and employs 1.1 million people. Yet dealing with plastic waste is hugely costly; removing litter, most of which is plastic, from the west coast of the United States costs taxpayers \$520 million each year. Also, the production of safer materials would spur innovation and boost employment in research and development. In fact, in the past three years or so, some plastics manufacturers themselves, under pressure from lobbyists and perhaps perceiving that current practices are unsustainable, have called for closed-loop systems.

If current consumption rates continue, the planet will hold another 33 billion tonnes of plastic by 2050. This would fill 2.75 billion refuse-collection trucks, which would wrap around the planet roughly 800 times if placed end to end. We estimate that this could be reduced to just 4 billion tonnes if the most problematic plastics are classified

as hazardous immediately and replaced with safer, reusable materials in the next decade. ■

**Chelsea M. Rochman** is in the School of Veterinary Medicine at the University of California, Davis, USA. **Mark Anthony Browne** is at the National Center for Ecological Analysis and Synthesis, Santa Barbara, California, USA. **Benjamin S. Halpern, Brian T. Hentschel, Eunha Hoh, Hrissi K. Karapanagioti, Lorena M. Rios-Mendoza, Hideshige Takada, Swee Teh, Richard C. Thompson.**

e-mails: [cmrochman@ucdavis.edu](mailto:cmrochman@ucdavis.edu);

[browne@nceas.ucsb.edu](mailto:browne@nceas.ucsb.edu)

1. Uhrin, A. V. & Schellinger, J. *Mar. Pollut. Bull.* **62**, 2605–2610 (2011).
2. Browne, M. A., Dissanayake, A., Galloway, T. S., Lowe, D. M. & Thompson, R. C. *Environ. Sci. Technol.* **42**, 5026–5031 (2008).
3. Lithner, D., Larsson, A. & Dave, G. *Sci. Total. Environ.* **409**, 3309–3324 (2011).
4. Teuten, E. L. et al. *Phil. Trans. R. Soc. B* **364**, 2027–2045 (2009).
5. Rochman, C. M., Hoh, E., Hentschel, B. T. & Kaye, S. *Environ. Sci. Technol.* <http://dx.doi.org/10.1021/es303700s> (2012).
6. Browne, M. A. et al. *Environ. Sci. Technol.* **45**, 9175–9179 (2011).
7. Pauly, J. L. et al. *Cancer Epidemiol. Biomarkers Prev.* **7**, 419–428 (1998).
8. Mettang, T. et al. *Nephrol. Dial. Transpl.* **11**, 2439–2443 (1996).
9. vom Saal, F. S. & Hughes, C. *Environ. Health Perspect.* **113**, 926–933 (2005).
10. Gaylor, M. O., Harvey, E. & Hale, R. C. *Chemosphere* **86**, 500–505 (2012).

Further reading accompanies this article online at [go.nature.com/p8sgip](http://go.nature.com/p8sgip).





The slender bow of the Infinity Bridge in Stockton, UK, is tied together through the deck to avoid the need for massive foundations.

## ECO-ENGINEERING

# Living in a materials world

From concrete to plastics, the megatonnes of stuff in the built environment are mostly manufactured and used with little thought for waste and pollution. Radical moves are afoot to refashion the urban fabric.

To the eco-engineer, the glass is neither half-full nor half-empty. It is simply twice as big as it needs to be. Building with maximum efficiency and minimal materials is increasingly urgent in our resource-strapped times. Many of today's structural engineers and designers are looking to natural forms and materials as the tried-and-tested guide.

The power and economy of evolved 'design' — eggshell, spiderwebs, bone — are inspiring architects to

**➔ NATURE.COM**  
For a design special on ecosanitation, see:  
[go.nature.com/d2hrxg](http://go.nature.com/d2hrxg)

experiment with solutions that work in harmony with physical forces and mimic biological form. Meanwhile, others are embracing 'extreme upcycling' in the flow of materials through the urban fabric, exploring ideas from edible upholstery to walls created from substances sourced in beetle exoskeletons.

Here, we look to the design ideas of three top players in the materials world: Olympic Velodrome engineer Chris Wise on lean, intelligent structures; architect and biomimicist Michael Pawlyn on a 3D-printed built environment; and chemist Michael Braungart on manufacturing beyond sustainability. ■

## CHRIS WISE

## Build with an eye on nature

*The Olympic Velodrome engineer on designing for strength and elegance.*

More than 300 years ago, Antonio Stradivari was making beautiful violins, each weighing a few hundred grams yet able to fill the largest concert hall with soul-lifting sound. Stradivari would never have said, "For safety, I'll double the thickness of the wood". If designed like most modern buildings, his violins would play like tree trunks. Stradivari's structures are balanced, natural systems: no imitations of birds' nests or deliberate distortions of geometry in the name of 'architecture' here.

Today's engineering is too full of 'tree-trunk' buildings: underdesigned, and over the whole planet. Astonishingly, the materials used in every three buildings designed in this lazy way could make at least four buildings — and with intelligence, even six.

Yet although humans have known about harmony between materials and structure for generations, every engineer, on every project, faces a mental battle with fear and hope.

Engineers are taught to design from fear, to avoid failure. The construction industry reinforces this, rewarding those who take the fewest risks, sacrificing our global material and energy stock on the altar of expediency. Dare we hope for wiser engineering, with beautiful performance from the least material?

Occasionally, special projects allow us to try. My firm Expedition Engineering was structural designer for the Velodrome in London's 2012 Olympic Park and the Infinity Bridge in Stockton, UK. Both won Britain's top prize in structural engineering, the Institution of Structural Engineers' Supreme Award for Engineering Excellence.

Both structures are 'form-found', shaped to be in equilibrium with the forces acting on them. Catalan architect Antoni Gaudí first popularized the technique. His Sagrada Família cathedral in Barcelona, Spain, begun in the 1880s, was effectively shaped upside-down — Gaudí's models were bags of sand hung from tension strings. Form-finding now uses digital-analysis engines for the behaviour of everything from cats' cradles and soap bubbles to giant basket-like grid shells. This way, structures can be sculpted to carry loads either in pure tension (like a spider's web) or in pure compression (like an eggshell).

Such structures embody the ancient Greek ideal of an inner beauty, carrying maximum load with minimum material in a way that cannot be bettered. Despite humanity's love of graceful curves and our need to use materials wisely, form-finding is still the exception. It should be the rule.

The Velodrome spans 130 metres with a tension roof structure only 76 millimetres thick; roofs of other stadia worldwide covering a comparable area are often metres deep. The Velodrome design achieved that lightness by letting nature lead, following the forces until they reached the equilibrium shape of a saddle in pure tension, anchored directly to the curved seating bowl. The forces are carried in harmony completely within the structural geometry, rather than outside it. It's an old trick much loved by the builders of Gothic cathedrals, although now we use a tension system of machine-woven steel cables, rather than a compression system of individually hand-cut stones.

If the Velodrome is a structure acting in tension, Stockton's double-arched Infinity Bridge is its complementary opposite. Through pure compression, the arches' shapes carry their own weight and the suspended deck: all the heavy forces are carried down the absolute centre of the structure. Because they are linked into a structure resembling an archer's bow, the two arches also act like a giant see-saw to resist the much lighter fluctuating weight of crossing pedestrians. (An adaptive bridge geometry, that changes continually in response to pedestrians' movement, could be coming soon.)

The ancient Roman Alcántara Bridge in Spain shows how far arches have evolved. The Romans did not know exactly where their forces went, so hedged their bets by infilling the massive individual masonry arches with cemented rubble to guarantee a pure compression line at least somewhere within the structure. Centuries later, the suspension bridge emerged. From Bristol's Clifton Suspension Bridge to San Francisco's Golden Gate, these are structures of great efficiency, but demand extraordinary tension anchorages buried at each end to hold the main suspension cables.

Infinity avoids such foundations by using tension cables in the deck to tie the ends of the arches together: the whole bridge just kisses the ground lightly at each end. The flow of the forces is written into the air in ultra-slender structural steel, rather than hidden inside approximately shaped stonework weighing thousands of tonnes.

Confident engineering comes from proper understanding of the natural phenomena to which it will be subject, and the more experimental, the more chance there is that something will catch us out. In the late 1990s, I was the engineering firm Arup's director for London's Millennium Bridge spanning the River Thames. It, too, was an off-piste, pure-tension natural structure. Too natural, perhaps: on its opening day, it wobbled harmonically like a giant guitar as the lateral sway of pedestrians' gaits became the vibration of the bridge. That wobble, however, forced a research project to find the cure: energy-absorbing dampers, choreographed by the late Arup engineer Tony Fitzpatrick and published for all to use.

The lessons learnt there fed into Infinity a decade later.

Despite these examples, and others from the likes of German architect and engineer Frei Otto, the late Peter Rice, and Tristram Carfrae at Arup, engineering suffers from a chronic sickness for which the construction industry is both cause and potential cure. 'Normal' fees for most engineering commissions are still based on a percentage of the construction cost: the more material you use, the more expensive the project, and the bigger your payout as an engineer.

Construction regulations are full of emotive words, such as 'collapse' (avoid it) and 'vibration' (get rid of it). Beyond these sanctions, they are largely silent. There is nothing in most commissions to encourage engineers to use less

material, so they don't. Yet if engineers were educated to design, say, perfectly tailored beams instead of off-the-shelf steel joists, we would cut about 30% from the millions of tonnes of steel beams used yearly. This should become the industry norm.

The huge construction supply chain also requires huge investment in new technologies. If manufacturers will not retool on a speculative basis, engineers need other research partners for innovative alternatives, such as those 'perfect' beams whose shape is tuned to the bending in them.

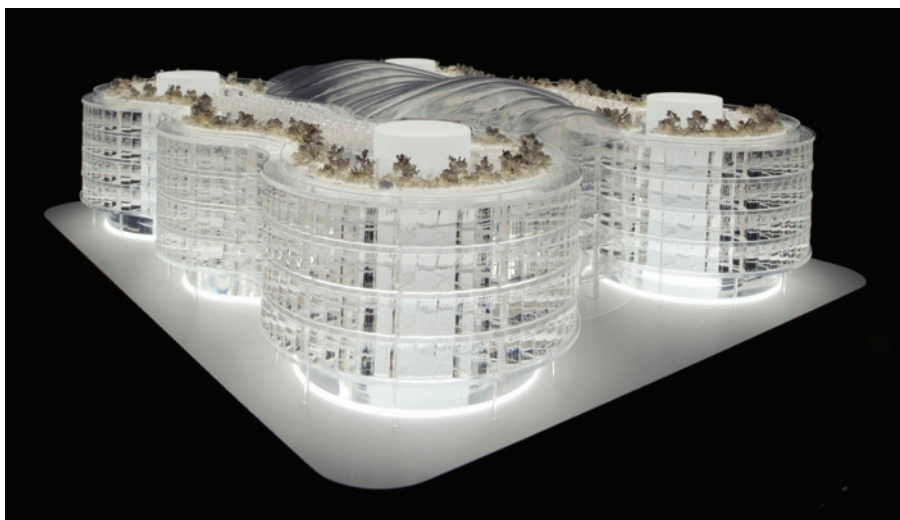
Infinity, the Velodrome, Otto's Olympic Stadium in Munich, Germany, and the Strad violin demonstrate what is possible. It may be ancient, this job of doing for a penny what any fool can do for a pound, but some of us seek performance through harmony between materials and natural forces. We design in hope, not fear — and with an eye on nature. ■

**Chris Wise** is a design engineer, co-founder of Expedition Engineering, Professor of Civil Engineering Design at University College London, and last year was awarded the gold medals of both the UK Structural and Civil Engineering Institutions for his design work. e-mail: [chris.w@expedition.uk.com](mailto:chris.w@expedition.uk.com)



London's 2012 Olympic Velodrome was built with the minimum of materials.





Exploration Architecture's biomimetic office building borrows from skulls and shells for form and function.

## MICHAEL PAWLYN Push the limits of 3D printing

*The biomimicist and architect on materials that borrow from nature.*

Biomimicry — the development of solutions based on biological adaptations — is one of the most exciting frontiers in design. After 3.8 billion years of natural research and development, organisms have solved many of the technical design problems that humans grapple with, but with a much greater economy of means. A key challenge for designers and architects is how to duplicate that finely realized engineering with the right materials.

Birds, for example, have evolved in response to particularly intense selective pressure on weight. The result is maximum strength, minimum mass. A magnified cross-section of a crow or magpie skull shows multiple bony shells connected by a matrix of ties and struts: an astounding combination of shell action (delivering material efficiency through its stiff shape) with space-frame technology (an inner structure subdivided for strength). Folding, vaulting and ribs in natural morphology can give organisms great structural efficiency.

But, as biomimetics expert Julian Vincent says, in nature “shape is cheap but material is expensive”. In architecture the opposite applies: manufacturing complex forms for a building is nearly always more costly than handling simpler, more monolithic elements.

A digital manufacturing revolution is dawning that could dramatically change that equation. Three-dimensional (3D) printing or ‘rapid prototyping’ emerged in the 1980s, allowing designers to turn a computer model

into a physical one accurately, with minimal labour, and at ambient temperatures and pressures. Architects have long used 3D printing to create design-development prototypes, and it is also used to manufacture custom-designed small objects. Now, the possibility of mass production beckons, thanks to three advances: 3D printers big enough to produce the larger elements needed, a greater range of printing materials, and the plummeting cost of the technology.

My architectural practice has been working on designs for a biomimetic office building that strive for the material efficiencies found in biology. The aim is to create floor slabs that get close to the substantially hollow structure of bird skulls, and glazing systems with extremely thin glass, curved to mimic shell forms, to provide the required stiffness.

Since 2009, several small buildings have been 3D printed from materials similar to concrete. But these lack the efficient microstructures found in biology, such as the fine latticework of glass sponges or the complex of voids in bone, and have limited potential for recycling. In the next few decades, it should be possible to use natural, endlessly recyclable polymers such as cellulose, which is found in wood, or chitin, a component of insect and shellfish exoskeletons. Materials made with 1% of the embodied energy of conventional ones, in structures ten times as efficient, could revolutionize manufacturing and construction.

Much of this may be beyond our current capabilities, but it is not the realm of fantasy: the natural world is living proof. ■

**Michael Pawlyn** is director of the London-based architectural practice *Exploration*, and a speaker for the ideas organization *TED*. His latest book is *Biomimicry in Architecture*.  
e-mail: [mwp@exploration-architecture.com](mailto:mwp@exploration-architecture.com)

## MICHAEL BRAUNGART

### Upcycle to eliminate waste

*The chemist recasts materials as ‘nutrients’ in an endless loop.*

Fundamentally rethinking materials as either harmlessly biodegradable or endlessly recyclable substances can transform them into something akin to nutrients nourishing the ‘metabolism’ of industry. Using these overarching principles of ‘cradle to cradle’ design and ‘upcycling’ (recycling to improve a material’s value), architect William McDonough and I developed guidelines and criteria for products to fit this frame. More than 400 products have been certified, from bricks and plastics to babies’ nappies and lighting.

Biodegradable, non-hazardous materials have a vast array of uses in products ranging from car tyres to textiles. When the products are worn out, the constituents are broken down to renew the raw materials. For example, the Climatex Lifecycle fabric produced by Swiss company Gessner in Wädenswil, and widely used on office furniture, is the first textile safe enough to eat. It is 100% recyclable and made with cellulose fibres, non-toxic dyes and wool from sheep that have been cared for ethically and responsibly. The water coming out of the plant is tested by the local ministry and is cleaner than when it went in. Textile scraps from the factory are sold to local farmers to use as mulch, helping to renew the soil.



Climatex 100% recyclable textiles are safe enough to eat.

Non-biodegradable products, such as glass bottles and television sets, can be recycled in safe, perpetual cycles, eliminating the need to use anything but recyclable raw materials. These metals, rare elements and approved synthetic materials do not physically degrade, and because they are not discarded will not come into direct contact with the environment.

Leasing and take-back systems (in which the manufacturer recycles the used product) are another way of making upcycling work, partly because they allow the use of higher-quality recyclable material that can take more wear. Companies already lease cars, but carpets — traditionally laden with chemicals such as chlorinated pigments, which persist in the environment and are often toxic — can also have many incarnations. The Dutch carpet manufacturer Desso, based in Waalwijk, makes polyvinyl chloride (PVC)-free carpets that improve air quality, because the patented fibres capture particulate matter from the atmosphere. Desso's take-back system even extends to PVC-free carpets that have been produced by other companies. In recycling, the carpet backing is reused and the yarn is rewoven into new carpets.

Raw materials and energy supplier Van Gansewinkel, headquartered in Eindhoven, the Netherlands, goes one step further. The company helped the Dutch mattress manufacturer Auping to design a take-back system for traditional mattresses as well as their own recyclable products. The steel in the mattresses is melted and reused; the foam is recycled as judo mats. Van Gansewinkel has also developed a continuous loop of high-quality recycling for office paper. It collects used paper from customers, sends it to the paper company Steinbeis in Stuttgart, Germany, for sustainable recycling, and with its partner Océ, based in Venlo, the Netherlands, supplies customers with the 100% recycled paper.

Nature does not respond to interdependence by seeking to minimize itself out of existence, but by growing and flourishing. Similarly, the key to generating a productive and sustainable economy is not through strategies of damage control and minimization, but through nourishing the industrial metabolism. ■

**Michael Braungart** holds chairs at Rotterdam School of Management and the University of Twente in the Netherlands, and is a founder of EPEA International Environmental Research in Hamburg, Germany. His latest book with William McDonough, *The Upcycle: Beyond Sustainability — Designing for Abundance*, will be published in April. e-mail: braungart@braungart.com

## THEATRE

# The needle in Newton's eye

A play about Isaac Newton's self-experimentation illuminates scientific rivalry, finds **Alla Katsnelson**.

Scientists occasionally conduct experiments on themselves. Among the most famous was Isaac Newton's extraordinary method for probing the nature of colour. He stuck a bodkin, a long sewing needle with a blunt point, into his eye socket, between eye and bone, and recorded seeing coloured circles and other visual phenomena. In his new play, *Isaac's Eye*, Lucas Hnath uses this bizarre experiment to explore scientific rivalry, the nature of truth and knowledge, and how the narratives of science and life congeal.

*Isaac's Eye* headlines the fifteenth annual First Light festival, a collaboration between the Alfred P. Sloan Foundation and the Ensemble Studio Theatre in New York to fund drama that explores scientific concepts and personalities. Hnath has looked to science for creative fodder since his undergraduate days at New York University. His first such work, which won a Sloan-sponsored writing competition, was a screenplay whose protagonist, computer scientist Adan Turner, finds himself imitating Alan Turing.

"I tend to write characters who try to push some kind of limit — who are trying to experience something that no one has ever before experienced," he says. "Inevitably, you have to deal with science if you go in that direction."

In his new drama, Hnath plays with history, presenting a fictional backstory to the legendary conflict between Newton, when he was still unknown, and the well-respected physicist Robert Hooke. "There's a law named after me," Hooke brags repeatedly throughout the play. In 1665, when the play is set, the plague is ravaging England while an ambitious Newton is being eaten away by the desire to join the Royal Society. Newton writes repeatedly to Hooke — then-Curator of Experiments — demanding that he be considered for membership. Receiving no reply, he sends Hooke a package containing the sole copy of all his writings. When Hooke sees that much of Newton's research treads the same ground as his own, he decides to visit Newton and take him down a peg or three. The

fictional encounter shakes up the course of both men's lives, as well as that of Catherine Storer, an apothecary's daughter who may have been a youthful romantic interest of Newton's.

The tension between the childish Newton, with his unpleasant moralistic streak allied to a tendency to fight dirty, and the callous, hedonistic Hooke, fuels the gripping narrative. Catherine, a realist, grounds the saga by trying to tempt both men to give up research for a mundane life. But perhaps the real story here is in how truth is told and perceived. Historians have criticized Newton for embellishing facts: his theory of gravity, for example, might not have been inspired by a falling apple. This tendency towards myth-making emerges early in the play, when Hooke accuses Newton of writing up experiments as though they were conducted in a single afternoon. Newton can't grasp the criticism. "It's clearer that way," he responds.

Sensing Hooke's reluctance to support him, Newton tells a lie. He claims that he has stabbed a bodkin under his eye to prove his own theory about the nature of light — and disprove Hooke's. Hooke calls his bluff, but when the scientists perform the experiment for real — first on a semi-willing subject, then on Newton — the truth is no clearer. The wincing in the audience during these highly realistic scenes was almost palpable, however.

"In some ways, this play is about how many liberties you decide to take when trying to convey something that's true," says Hnath. To help viewers sort fact from fiction, Hnath employed the conceit of a nameless narrator who uses cardboard signs and a chalkboard to wrangle known facts about the two scientists' lives into a list of bullet points. It's the seventeenth-century version of PowerPoint. But in the end, tweaking reality — as Hnath so skillfully does — might provide the clearer picture of the human truths in this scientific tug-of-war. ■

**Alla Katsnelson** is a freelance writer living in New York.  
e-mail: akatsnelson@gmail.com

*Perhaps the real story is in how truth is told and perceived.*



# Correspondence

## Resistant TB: use the tools available

Rwanda's experience in combating multi-drug-resistant tuberculosis (MDR-TB) offers lessons in using available tools wisely now, rather than waiting for the results of long-term scientific investment (*Nature* **493**, 14–16; 2013).

Of 306 patients who began standardized 'second-line' drug regimens for MDR-TB through a Rwandan programme in 2005–09, 270 (88.2%) were successfully treated. Nearly half of the patients were HIV-positive and all lived in extreme poverty.

Crucially, Rwanda's Ministry of Health provided each patient with close follow-up care and with monthly support towards nutrition and transport. As a result, only six patients (2%) defaulted on their treatment.

Such common-sense interventions prevent disease progression and facilitate directly observed therapy. But they are seldom included in cost-effectiveness studies — much less in nationwide strategies.

Once drug-supply chains and comprehensive support for adherence to treatment are in place, such positive MDR-TB treatment outcomes will be possible all over rural Africa.  
**Agnes Binagwaho\*** *Ministry of Health, Kigali, Rwanda.*  
[agnes\\_binagwaho@hms.harvard.edu](mailto:agnes_binagwaho@hms.harvard.edu)

*\*On behalf of 5 co-signatories (see [go.nature.com/ttsddo](http://go.nature.com/ttsddo) for full list).*

## Resistant TB: remove cultural barriers

Social and cultural obstacles confront practitioners trying to control the spread of multi-drug-resistant tuberculosis (TB; *Nature* **493**, 14–16; 2013). Removing these barriers will improve adherence to drug regimes and impede the development of drug resistance.

For example, young Asian women with TB may not come

forward for treatment because they see it as a threat to their marriage prospects. Alcohol consumption, which is prevalent among some Caucasian patients, can interfere with drug treatments. And recovering patients who return to their jobs, often away from home, may drop contact with health-care providers.

Besides supplying medicines and efficient health-service management, it is vital to promote culturally adjusted messages — in particular, to emphasize why patients with TB must stick to their lengthy treatments.

While waiting for shorter and less-toxic drug regimens to become available on a large scale, non-governmental organizations such as Médecins Sans Frontières include socio-cultural support to help promote adherence. This strategy is founded on medical anthropology — the applied study of traditional beliefs and practices relating to illness and health care. Such support helps to ensure effective treatment and follow-up (see, for example, [go.nature.com/4b6f9f](http://go.nature.com/4b6f9f)).

**Gilles Guerrier Noumea,**  
*New Caledonia.*  
[guerriergilles@gmail.com](mailto:guerriergilles@gmail.com)

## Altmetrics: research council responds

Heather Piwowar lists several funders, including the UK Medical Research Council (MRC), that “still explicitly ask for a list of research papers rather than products” in their grant-application process (*Nature* **493**, 159; 2013). However, we at the MRC strongly agree that other outputs can provide valuable evidence of research significance and impact.

These research ‘products’ include development of new interventions, production and distribution of widely used data sets, changes to clinical practice or other advances that influence policy decisions, and commercial indicators such as

licence income and growth of spin-out companies.

The MRC considers all such products in assessing researcher track records. Our grant-application forms specifically ask researchers how they intend to manage and share the results of their work, and to outline their productivity beyond published papers.

In addition, the company Researchfish has developed a structured online system for MRC-funded scientists to record their publications and research-product output. The system also provides publicly accessible information relating to MRC grants. It has now been adopted by 15 other UK funding agencies.  
**Ian Viney** *Medical Research Council, London.*  
[ian.viney@headoffice.mrc.ac.uk](mailto:ian.viney@headoffice.mrc.ac.uk)

## Altmetrics: too soon for use in assessment

I question the worth of many of the altmetrics (‘alternative metrics’) that Heather Piwowar discusses for evaluating a researcher's total output (*Nature* **493**, 159; 2013).

Altmetrics include, for example, the number of downloads, ‘likes’ or shares for a research-related YouTube video, tweet or blog post. Although these may provide insight into how such research ‘products’ have influenced the community and the public, they lack authority and credibility as a performance measure, not least because it is easy to cheat by creating multiple accounts.

Grant reviewers cannot therefore rely on such data as indicators of the value of their associated research products. In fact, the US National Science Foundation's new funding policy, around which Piwowar bases her arguments, requires the research products to be in citable and accessible form — implying that it is only the number of citations in scientific journals that will be taken into account in assessing

the value of these products.

I believe that it is premature to integrate most altmetrics, apart from citation statistics, into research-assessment schemes at a time when the merits of publication quantity versus quality are still being debated (N. Haslam and S. M. Laham *Eur. J. Soc. Psychol.* **40**, 216–220; 2010).

**Man Kit Cheung** *The Chinese University of Hong Kong, Hong Kong.*  
[mkcheung@cuhk.edu.hk](mailto:mkcheung@cuhk.edu.hk)

## Aristocratic names get short shrift

In his discussion of suspects in the Piltdown Man mystery, Chris Stringer alludes to the French Jesuit priest, philosopher and palaeontologist Pierre Teilhard de Chardin (*Nature* **492**, 177–179; 2012). The article mistakenly shortens the great man's compound surname to “de Chardin”. In fact, it should read “Teilhard de Chardin” or just “Teilhard”.

Such aristocratic last names are often misinterpreted in English. One glaring example is the condition known as Tourette's syndrome, incompletely named after Georges Gilles de la Tourette, who first described it. “Gilles” is part of the compound last name and not a middle name, as is commonly assumed.

**Antoine Louchart** *Ecole Normale Supérieure de Lyon, Lyons, France.*  
[antoine.louchart@ens-lyon.fr](mailto:antoine.louchart@ens-lyon.fr)

### CORRECTION

A table in the Outlook article ‘Nanotechnology: Carrying drugs’ (*Nature* **491** (suppl. 7425), S58–S60; 2012) wrongly stated that Cerulean Pharma's drug CRLX101 is in phase I trials. In fact, it has been in phase II trials since mid-2012. The table also incorrectly gave the company name as “Cerulean Pharmaceuticals”.

## To go or not to go

A study shows that, rather than sequentially starting and stopping a movement, two parallel pathways involving neurons in the brain's basal ganglia seem to work in tandem to accomplish the complex task of motion. [SEE LETTER P.238](#)

D. JAMES SURMEIER

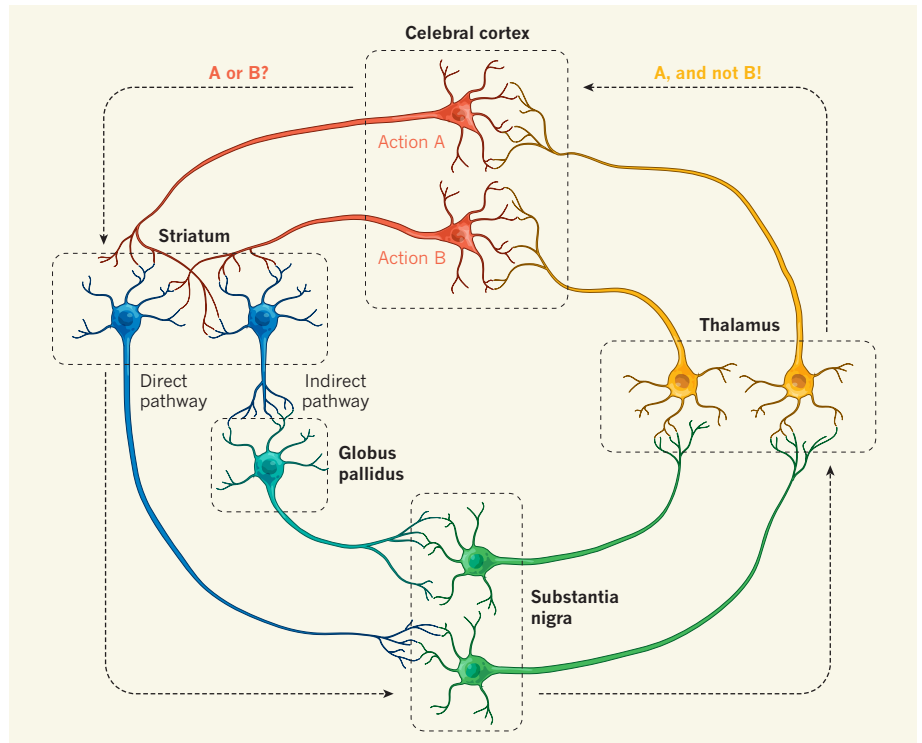
Think of the basal ganglia as a trusted adviser. This richly interconnected collection of neuronal clusters in the brain's subcortical region helps us to decide what to do when we are presented with a choice. The largest of these clusters — the striatum — stores information about past experiences. When interrogated by another brain region, the cerebral cortex, about what might be the most rewarding course of action, the striatum retrieves the information on past experience and passes a recommendation through the thalamus back to the cerebral cortex, which makes the decision. It has generally been thought that the striatal recommendation is a binary 'yea' or 'nay'. An elegant study by Cui *et al.*<sup>1</sup> on page 238 of this issue challenges this model and suggests that the striatum makes a recommendation on both what to do and what not to do\*.

A good metaphor for the basal ganglia has been a car. To go, you take your foot off the brake and press down on the accelerator; to stop, you do the opposite. Similarly, the basal ganglia are organized into two parallel pathways, which are anchored by spiny projection neurons (SPNs). It is thought that the SPNs in the direct pathway (think accelerator) promote action, whereas those in the indirect pathway (brake) suppress it.

Much of the evidence for this model comes from clinical observations of patients with Parkinson's disease<sup>2</sup>. Loss of the neurotransmitter dopamine in these individuals makes the direct-pathway SPNs sluggish and the striatal, indirect-pathway SPNs jumpy. Consequently, the indirect-pathway SPNs are likely to win any discussion about what to do, resulting in a constant nay recommendation — as if the brake were constantly being pressed firmly down. This is thought to lead to the difficulty in initiating movement that is characteristic of Parkinson's disease.

Several observations have been difficult to reconcile with this model<sup>3–5</sup>. However, studies using optogenetic tools to selectively activate direct or indirect pathways, or deletion of key signalling molecules, have reinforced<sup>6,7</sup> the notion that these two networks behave

\*This article and the paper under discussion<sup>1</sup> were published online on 23 January 2013.



**Figure 1 | Decision-making at the neuronal level.** In this simplified model, cerebral cortex neurons (red) ask striatal neurons (blue) for advice on two possible courses of action: action A, which has previously had a positive outcome, and action B, which has had a negative outcome. It had been thought that, in response, two groups of striatal neurons would activate substantia nigra neurons (green) through either a direct pathway favouring action A or an indirect pathway, with an intermediate step in the globus pallidus, advising against action B. Cui *et al.*<sup>1</sup> report, however, that both direct and indirect pathways are activated simultaneously to recommend action A and advise against action B. These recommendations are then passed through neurons in the thalamus (yellow) to the cerebral cortex.

like on and off switches for movement.

Surprisingly, there have been no investigations of what each SPN group does when an animal is asked to make a choice or initiate a movement, to see if the model is correct. This is because direct- and indirect-pathway SPNs are intermingled and difficult to tell apart using physiological tools. Cui *et al.* have developed an ingenious strategy to overcome this obstacle. They used genetically engineered mice that selectively express an optical indicator of changes in intracellular calcium concentration in just one or the other type of SPN. The indicator acted as a 'reporter', providing an optical signal when an SPN became active. Next, the authors placed tiny fibre-optic cables

at precise locations in the striatum; the animals were then trained to press a lever for a food reward, while the activity of their direct- and indirect-pathway SPNs was monitored.

The classical model would predict that, when a mouse began to move, direct-pathway SPNs would become active, with indirect-pathway SPNs becoming active when it stopped. This is not what Cui and colleagues saw. Instead, they found that both groups of SPNs became active before the initiation of movement, as if both were making recommendations about what to do (Fig. 1). And when the mouse stopped moving, both groups of SPNs became relatively inactive.

That direct- and indirect-pathway SPNs do



not simply make sequential yes or no recommendations kills the car metaphor. It is possible, albeit unlikely, that the activity of these neurons is somehow combined to make a binary decision, but Cui and colleagues found no indication of this. It is more likely that the striatum is simultaneously making recommendations about what to do and what not to do. Intuitively, this is appealing. When choosing between two alternative courses of action, it is always best to have a reason for choosing one and for not choosing the other.

This model also fits with current knowledge of synaptic plasticity in the striatum<sup>8</sup>. (This phenomenon describes activity-dependent changes in the strength of synaptic connections between neurons.) When we do something that turns out well, a pulse of dopamine is released in the striatum. This strengthens the synaptic connections between cortical neurons and direct-pathway SPNs, making them more likely to respond to these cortical neurons in the future. Assuming that these cortical neurons represent a particular action, when we are next faced with a decision in which this action is possible, they become active. Because of the strengthened connections, direct-pathway SPNs will tend to become active too and, on the basis of the current model, would send a message back to the cortex indicating that this is a good course of action.

An opposing scenario — when a course of action has a bad outcome — is played out in the indirect-pathway SPNs. A drop in dopamine release in the striatum promotes the strengthening of connections between cortical neurons and these SPNs. So, in the future, when the cortical neurons representing these actions become active, indirect-pathway SPNs follow suit, again theoretically, sending back a warning message. This implies that striatal recommendations are about specific movements and are not generalized stop or go commands.

One intriguing question is whether ensembles of direct- and indirect-pathway SPNs exist that are specific for particular movements. Cui *et al.* could not resolve this level of detail. Another question of importance for translating the results into clinical applications is how the basal-ganglia neurons sitting between the striatum and the thalamus could pass on more than one recommendation. For this, the neurons would have to work independently, with some recommending one action and others recommending against a conflicting action. A recent study<sup>9</sup> suggests that these neurons show the sort of dynamic coupling to direct and indirect pathways that would allow this to happen.

In primate models of Parkinson's disease, neurons of the globus pallidus and substantia nigra that forward the recommendations from the striatum seem to be forced into 'group-thinking', spiking in synchrony<sup>10</sup>. Similar behaviour is seen in the neurons of patients with Parkinson's disease<sup>11</sup>. This could be interpreted by the thalamus and cortex

as a recommendation to act in two mutually exclusive ways at the same time, which would lead to difficulty in making a decision and initiating movement — a cardinal feature of the disease. Therapies designed to allow these neuronal clusters to pass on the recommendations of both direct- and indirect-pathway SPNs, as seen by Cui *et al.* for movement initiation, might restore the ability of patients with Parkinson's disease to move more readily. ■

**D. James Surmeier** is in the Department of Physiology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA.

e-mail: j-surmeier@northwestern.edu

#### ASYMMETRIC SYNTHESIS

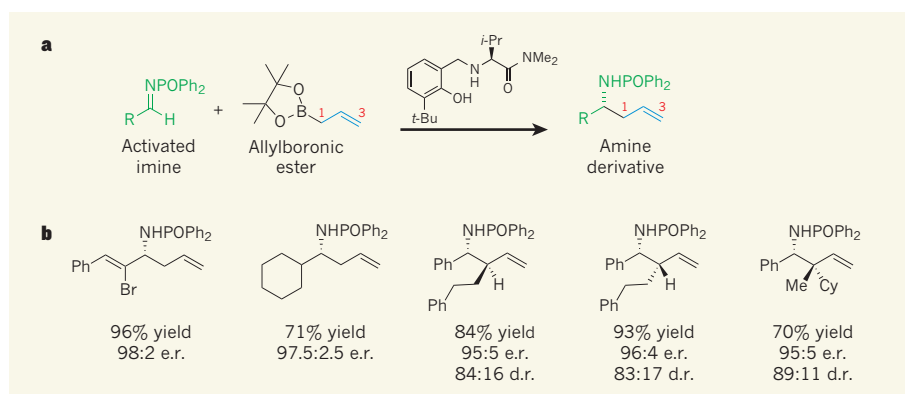
## Relay catalysis at a boron centre

A boron complex catalyses the addition of allyl groups — hydrocarbon motifs — to 'activated imines' in a relay-like process, generating synthetically useful compounds as single mirror-image isomers. [SEE LETTER P.216](#)

VALER JESO & GLENN C. MICALIZIO

The significance of chirality — the property of 'handedness' — in nature cannot be understated. The characteristic is inherent in the molecules that define life, and in most of the biologically active, naturally occurring compounds that have served as a foundation for drug discovery over the past century<sup>1</sup>. Not surprisingly, therefore,

chemists have long sought asymmetric methods to convert non-chiral starting materials to single enantiomers (isomeric products of a single handedness). On page 216 of this issue, Silverio *et al.*<sup>2</sup> report a great conceptual advance towards this goal. The authors describe a catalyst for an important class of carbon-carbon bond-forming reactions that enables the enantioselective addition of allyl groups (allyl is CH<sub>2</sub>CH=CH<sub>2</sub>) to activated



**Figure 1 | An asymmetric allyl-transfer reaction.** **a**, Silverio *et al.*<sup>2</sup> report a catalytic reaction in which an allyl group (blue) is transferred to an activated imine to form an amine derivative as a single enantiomer. Carbon atoms at either end of the allyl group are numbered. Me, methyl (CH<sub>3</sub>); Ph, phenyl (C<sub>6</sub>H<sub>5</sub>); *t*-Bu, tert-butyl (CMe<sub>3</sub>); *i*-Pr, isopropyl (CHMe<sub>2</sub>); R can be a variety of saturated and unsaturated hydrocarbons. **b**, Some of the chiral products prepared by Silverio *et al.*, the percentage yields obtained and the enantiomeric ratios (e.r., the ratio of the amount of depicted isomer to that of the mirror-image isomer produced in each reaction). For those compounds that can form as isomers known as diastereomers, the diastereomeric ratio (d.r., the ratio of the amount of the depicted diastereomer to that of the alternative diastereomer) is also shown. Cy is a cyclohexyl group, C<sub>6</sub>H<sub>12</sub>.

not simply make sequential yes or no recommendations kills the car metaphor. It is possible, albeit unlikely, that the activity of these neurons is somehow combined to make a binary decision, but Cui and colleagues found no indication of this. It is more likely that the striatum is simultaneously making recommendations about what to do and what not to do. Intuitively, this is appealing. When choosing between two alternative courses of action, it is always best to have a reason for choosing one and for not choosing the other.

This model also fits with current knowledge of synaptic plasticity in the striatum<sup>8</sup>. (This phenomenon describes activity-dependent changes in the strength of synaptic connections between neurons.) When we do something that turns out well, a pulse of dopamine is released in the striatum. This strengthens the synaptic connections between cortical neurons and direct-pathway SPNs, making them more likely to respond to these cortical neurons in the future. Assuming that these cortical neurons represent a particular action, when we are next faced with a decision in which this action is possible, they become active. Because of the strengthened connections, direct-pathway SPNs will tend to become active too and, on the basis of the current model, would send a message back to the cortex indicating that this is a good course of action.

An opposing scenario — when a course of action has a bad outcome — is played out in the indirect-pathway SPNs. A drop in dopamine release in the striatum promotes the strengthening of connections between cortical neurons and these SPNs. So, in the future, when the cortical neurons representing these actions become active, indirect-pathway SPNs follow suit, again theoretically, sending back a warning message. This implies that striatal recommendations are about specific movements and are not generalized stop or go commands.

One intriguing question is whether ensembles of direct- and indirect-pathway SPNs exist that are specific for particular movements. Cui *et al.* could not resolve this level of detail. Another question of importance for translating the results into clinical applications is how the basal-ganglia neurons sitting between the striatum and the thalamus could pass on more than one recommendation. For this, the neurons would have to work independently, with some recommending one action and others recommending against a conflicting action. A recent study<sup>9</sup> suggests that these neurons show the sort of dynamic coupling to direct and indirect pathways that would allow this to happen.

In primate models of Parkinson's disease, neurons of the globus pallidus and substantia nigra that forward the recommendations from the striatum seem to be forced into 'group-thinking', spiking in synchrony<sup>10</sup>. Similar behaviour is seen in the neurons of patients with Parkinson's disease<sup>11</sup>. This could be interpreted by the thalamus and cortex

as a recommendation to act in two mutually exclusive ways at the same time, which would lead to difficulty in making a decision and initiating movement — a cardinal feature of the disease. Therapies designed to allow these neuronal clusters to pass on the recommendations of both direct- and indirect-pathway SPNs, as seen by Cui *et al.* for movement initiation, might restore the ability of patients with Parkinson's disease to move more readily. ■

**D. James Surmeier** is in the Department of Physiology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA.

e-mail: j-surmeier@northwestern.edu

## ASYMMETRIC SYNTHESIS

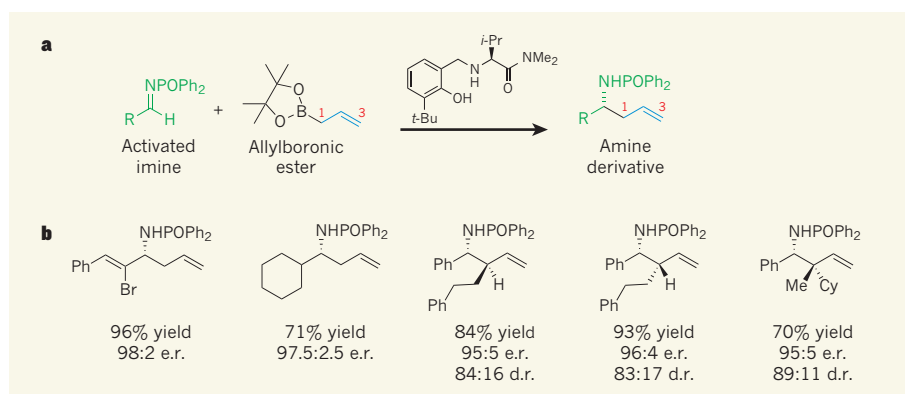
# Relay catalysis at a boron centre

A boron complex catalyses the addition of allyl groups — hydrocarbon motifs — to 'activated imines' in a relay-like process, generating synthetically useful compounds as single mirror-image isomers. [SEE LETTER P.216](#)

VALER JESO & GLENN C. MICALIZIO

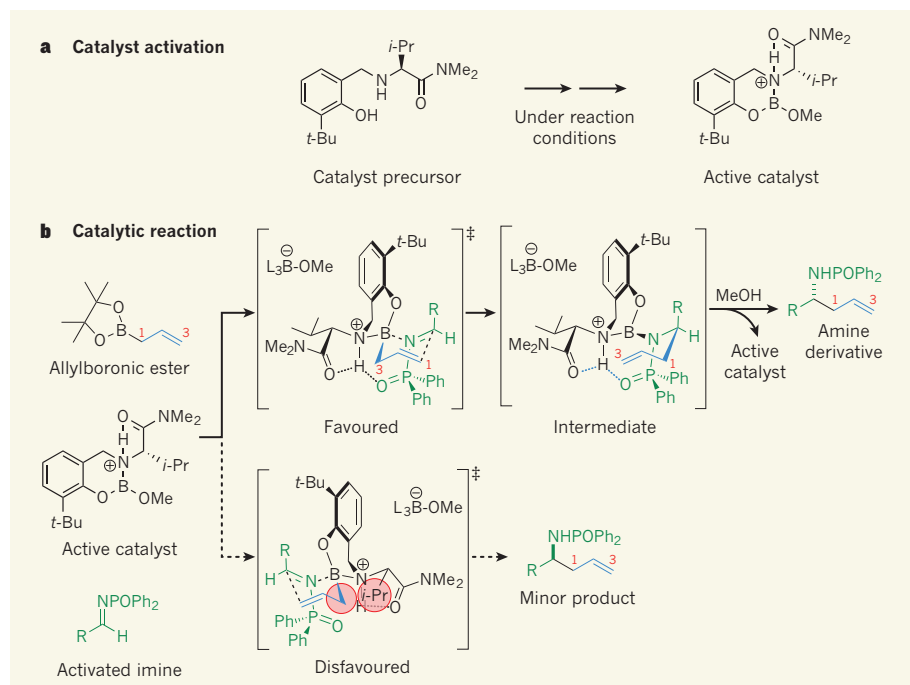
The significance of chirality — the property of 'handedness' — in nature cannot be understated. The characteristic is inherent in the molecules that define life, and in most of the biologically active, naturally occurring compounds that have served as a foundation for drug discovery over the past century<sup>1</sup>. Not surprisingly, therefore,

chemists have long sought asymmetric methods to convert non-chiral starting materials to single enantiomers (isomeric products of a single handedness). On page 216 of this issue, Silverio *et al.*<sup>2</sup> report a great conceptual advance towards this goal. The authors describe a catalyst for an important class of carbon-carbon bond-forming reactions that enables the enantioselective addition of allyl groups (allyl is CH<sub>2</sub>CH=CH<sub>2</sub>) to activated



**Figure 1 | An asymmetric allyl-transfer reaction.** **a**, Silverio *et al.*<sup>2</sup> report a catalytic reaction in which an allyl group (blue) is transferred to an activated imine to form an amine derivative as a single enantiomer. Carbon atoms at either end of the allyl group are numbered. Me, methyl (CH<sub>3</sub>); Ph, phenyl (C<sub>6</sub>H<sub>5</sub>); *t*-Bu, tert-butyl (CMe<sub>3</sub>); *i*-Pr, isopropyl (CHMe<sub>2</sub>); R can be a variety of saturated and unsaturated hydrocarbons. **b**, Some of the chiral products prepared by Silverio *et al.*, the percentage yields obtained and the enantiomeric ratios (e.r., the ratio of the amount of depicted isomer to that of the mirror-image isomer produced in each reaction). For those compounds that can form as isomers known as diastereomers, the diastereomeric ratio (d.r., the ratio of the amount of the depicted diastereomer to that of the alternative diastereomer) is also shown. Cy is a cyclohexyl group, C<sub>6</sub>H<sub>11</sub>.





**Figure 2 | Catalytic cycle.** **a**, Silverio *et al.* propose that, in their allyl-transfer reactions<sup>2</sup>, the catalyst precursor is converted into the active catalyst *in situ*. **b**, The allylboronic ester, catalyst and activated imine can then form two possible transition states. One of these is disfavoured because the *i*-Pr group in the catalyst clashes (red circles) with part of the allyl group. The reaction therefore proceeds preferentially through the favoured transition state, forming an intermediate that reacts with one of the solvents in the mixture (methanol, MeOH) to provide the amine product and regenerate the active catalyst. Note that in the favoured transition state, the allyl group is bound to the boron atom at carbon 3, but then forms a C–C bond with the activated imine at carbon 1. Dotted lines indicate hydrogen bonds; dashed lines in the transition states indicate bonds that are forming; ‡, transition state;  $[L_3B-OMe]^-$  is an anion that forms when the allylboronic ester reacts with methanol.

imines ( $C=NPh$ ; Ph, phenyl group) or to electron-deficient carbonyls ( $C=O$ ). The reactions deliver homoallylic amine derivatives from activated imines, or homoallylic alcohols from carbonyls, and can establish molecular complexity far beyond that associated with transfer of a simple allyl unit.

Asymmetric methods aim to control the absolute stereochemistry of reaction products — the spatial arrangement of atoms that distinguishes an enantiomer from its mirror image. Such methods for adding allyl groups to carbonyls have a rich history, and have transformed the way in which many biologically active natural products are prepared in the laboratory<sup>3</sup>. Many of the widely adopted, practically useful protocols<sup>4</sup> are stoichiometric processes in which chirality is transferred to the product from a chiral reagent or auxiliary (a group that is temporarily attached to a substrate to control the enantioselectivity of a reaction). Alternatively, recent methods have been described in which absolute stereochemistry is derived from the transposition of chirality from the allyl donor<sup>5</sup>. Despite these advances, the majority of effective allylation processes use chiral auxiliaries and require substantial effort to purify the desired product from the reaction mixture.

In contrast to these modern synthetic

approaches, nature typically embraces non-stoichiometric (that is, catalytic) approaches to establish chirality in molecules. Although advances in the catalysis of allyl-transfer reactions<sup>6–8</sup> have begun to chip away at the challenging problem of emulating nature's masterful approach to inducing asymmetry, Silverio and colleagues' contribution marks a substantial development with the potential to be used widely in organic chemistry.

The authors have designed a small, chiral catalyst that is composed of a derivative of the amino acid valine and an electron-deficient boron atom, which orchestrates a highly enantioselective carbon–carbon (C–C) bond-forming reaction between an allyl donor (an allylboronic ester) and an activated imine (Fig. 1a). The researchers propose that the stereochemistry of the valine-derived moiety is catalytically translated to the product by way of a highly organized transition state that is supported by several key hydrogen bonds (in a manner that is somewhat enzyme-like). Notably, the authors have demonstrated that the catalytic process is useful for transformations that are more complex than simple allyl transfer (Fig. 1b).

Unlike many modern advances in catalysis, Silverio and co-workers' enantioselective reaction is quite straightforward to achieve. It takes

place in a mixture of commonly used solvents at room temperature, is complete in less than six hours and requires a tiny amount of the inexpensive catalyst precursor (0.25 mole per cent). The reaction will surely be useful for the synthesis of chiral amines, which are ubiquitous structural motifs in biology and of great interest to medicinal chemists.

The conceptual novelty of Silverio and colleagues' advance is firmly rooted in the manner in which the C–C bond-forming reaction proceeds, the factors that induce asymmetry (the formation of chiral centres) and the reaction's exceptional enantioselectivity and regioselectivity (the selectivity with which reactions occur at specific sites in molecules). To explain in more detail, let us consider reactions in which the allyl group from an allylboronic ester adds to an activated imine to form a chiral homoallylic amine derivative. The authors propose that a chiral catalyst precursor is initially converted to the active catalyst *in situ* (Fig. 2a) and then reacts with the non-chiral allylboronic ester, transferring the allyl group into a chiral environment. The resulting chiral, electron-deficient boron species can form one of two transition-state structures with the activated imine (Fig. 2b) — complexes defined by a network of hydrogen bonds between the valine-derived unit and substrate, and in which the allylboronic ester and electron-deficient imine mutually activate each other for the ensuing C–C bond-forming reaction. Silverio *et al.* suggest that one of the two transition-state structures is substantially destabilized by interactions between the allyl unit and a bulky group in the catalyst. The reaction therefore progresses selectively through the other transition state, causing the homoallylic amine derivative to form with high enantioselectivity.

The regioselectivity of this transformation is distinct from established allyl-transfer chemistry that involves allylboronic esters because the C–C bond formation occurs at the allyl carbon atom that was originally bound to the boron atom in the allylboronic ester, rather than at the distal carbon located at the terminus of the allyl group. This selectivity is a consequence of two regioselective allyl-transfer processes that occur in the catalytic mechanism, which is an interesting feature of the current asymmetric process.

Silverio and colleagues' advance marks a new direction in a growing field that capitalizes on the catalytic properties of amino-acid-metalloid complexes; metalloids are elements, such as boron and silicon, that are difficult to classify as distinctly metallic or non-metallic. In contrast to earlier work, which targeted catalysts for the asymmetric reduction of carbonyls and for cycloaddition reactions<sup>9,10</sup>, the authors' findings reveal a unique mechanistic pathway that establishes a distinct asymmetric transformation. We look forward with great enthusiasm to the development of other catalytic, asymmetric C–C

bond-forming processes that take advantage of the mechanistic features of this pathway. ■

**Valer Jeso and Glenn C. Micalizio** are in the Department of Chemistry, Scripps Research Institute, Scripps Florida, Jupiter, Florida 33458, USA.

e-mail: micalizio@scripps.edu

- Schreiber, S. L. *Nature Chem. Biol.* **1**, 64–66 (2005).
- Silverio, D. L. *et al. Nature* **494**, 216–221 (2013).
- Chemler, S. R. & Roush, W. R. in *Modern Carbonyl*

- Chemistry* (ed. Otera, J.) 403–490 (Wiley-VCH, 2001).
- Fu, J. & Denmark, S. E. *Chem. Rev.* **103**, 2763–2793 (2003).
- Chen, M. Z. J. *Org. Chem.* **75**, 8048–8059 (2010).
- Moran, J. & Krische, M. J. *Pure Appl. Chem.* **84**, 1729–1739 (2012).
- Denmark, S. E. & Almstead, N. G. in *Modern Carbonyl Chemistry* (ed. Otera, J.) 300–401 (Wiley-VCH, 2001).
- Lou, S., Moquist, P. N. & Schaus, S. E. *J. Am. Chem. Soc.* **129**, 15398–15404 (2007).
- Corey, E. J., Raman, K. B. & Shibata, S. *J. Am. Chem. Soc.* **109**, 5551–5553 (1987).
- Corey, E. J., Shibata, T. & Lee, T. W. *J. Am. Chem. Soc.* **124**, 3808–3809 (2002).

## CELL BIOLOGY

# Beneficial lessons from viruses

To thwart the antimicrobial responses of their hosts, pathogens have evolved diverse mechanisms, including autophagy. Knowledge of such mechanisms has now led to a pro-autophagy peptide that may be of therapeutic value. [SEE ARTICLE P.201](#)

ADOLFO GARCÍA-SASTRE

Autophagy is self-cannibalism at the cellular level; it results in the degradation of cytoplasmic components by cellular organelles called lysosomes. The aim of this process is either the rapid recycling of cellular material during starvation conditions or the disposal of damaged organelles and harmful proteins. Autophagy is also a notable pathway mediating innate immune responses that eliminate viruses and other intracellular microbes. It is therefore hoped that this process might be harnessed as a strategy for fighting microbial infections. To this end, on page 201 of this issue, Shoji-Kawata *et al.*<sup>1</sup> report the identification of a peptide that induces autophagy and

that has potent inhibitory properties against various disease-causing agents. Along the way, the authors have also discovered a previously unknown negative regulator of autophagy — the GPR-1 protein\*.

In the past decade, the understanding of antiviral innate immunity has become an intense area of research. This is because its complex regulation and interactions with other cellular pathways suggest that it may form the basis of crucial discoveries in cell biology. Moreover, the potent inhibitory effects of innate immune pathways on viral replication may allow these pathways to be exploited for the development of broad-spectrum antiviral

\*This article and the paper under discussion<sup>1</sup> were published online on 30 January 2013.

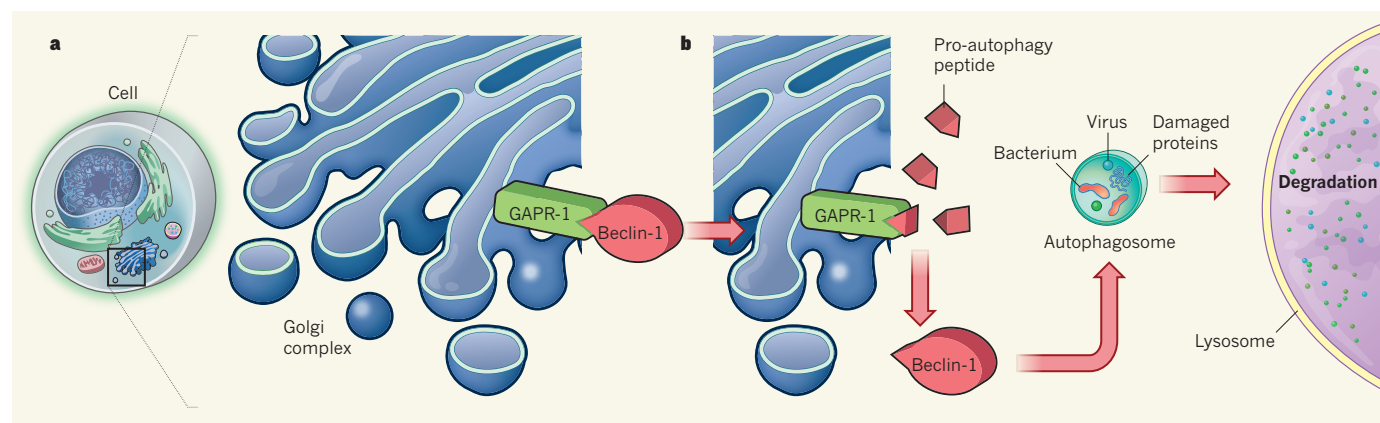
drugs to potentiate innate immunity. Indeed, the discovery of such antiviral innate immune responses mediated by interferon proteins resulted in the development of recombinant interferon as an antiviral drug that is used to treat infections caused by hepatitis C and certain other viruses<sup>2</sup>.

The importance of autophagy in the cellular antiviral response is illustrated by the ability of diverse viruses — including HIV-1, influenza virus and herpes viruses — to inhibit autophagy<sup>3–6</sup>. Understanding viral inhibition of autophagy may therefore lead to new ways to restore this process during viral infection and so eliminate the virus.

Shoji-Kawata and colleagues set out to explore the inhibition of autophagy by HIV-1. It is known<sup>3</sup> that, during infection, the Nef protein of HIV-1 inhibits autophagy by interacting with beclin-1 — a key positive regulator of autophagy that is required for the formation of autophagosomes. (Autophagosomes are double-membrane-bound vesicles that engulf cytoplasmic material and disease-associated agents for transport to lysosomes.)

The authors' detailed mapping of the beclin-1 domain that interacts with Nef revealed that a sequence of 18 amino-acid residues in beclin-1 is required for this interaction. Shoji-Kawata *et al.* next investigated the effects of delivering this 18-residue peptide to the cell, by fusing it with a cell-permeable peptide. When incubated with various kinds of cells, this short fusion peptide potently induced autophagy in all cell lines tested.

The authors also report that GPR-1, a cellular protein that associates with an intracellular complex called the Golgi, under normal circumstances sequesters beclin-1 to inhibit its autophagic properties. Intriguingly, the 18-residue peptide of beclin-1 could also interact with GPR-1, competing with, and releasing, beclin-1 from the Golgi into the cytoplasm, where it mediates the formation of autophagosomes (Fig. 1). Future studies



**Figure 1 | A promising peptide.** **a**, The Nef protein of HIV-1 (not shown) and the cellular protein GPR-1 interact with the same domain of the beclin-1 protein to inhibit the degradative process of autophagy. GPR-1 does so by sequestering beclin-1 in the cellular compartment known as the Golgi complex. **b**, Shoji-Kawata *et al.*<sup>1</sup> report that a peptide

derived from the same domain of beclin-1 interacts with GPR-1. This results in the release of beclin-1, allowing it to mediate the formation of autophagosomes, which engulf intracellular pathogens and harmful proteins to transport them to the lysosomal compartment for degradation.





## 50 Years Ago

*Animal Dispersion in Relation to Social Behaviour.* By Prof. V. C. Wynne-Edwards — The theme of this book is that over-exploitation of food resources by an animal population will lead to dissipation of the resources and deterioration of the population; that there exists an optimum level of any food resource for allowing sustained productivity; that this needs to be matched by the maintenance of an optimum level of the animal population exploiting it; and that this optimum is attained by means of the many behaviour patterns and devices (especially social ones) found among animals. The author further postulates that this end-result of evolution has been brought about in the past not by individual natural selection, but by group selection of populations ... The enormous weakness of this enormous book, so full of fascinating information, so impregnated with one particular teleological bias, is that it gives no single case-history of group selection ... Whether it occurs widely at all the courageous reader had better decide for himself.  
From *Nature* 16 February 1963

## 100 Years Ago

At the end of last month the president and fellows of Harvard College voted to establish the Harvard University Press, for the publication of works of a high scholarly character ... The function of a university press should be to publish works of prime importance and distinctive merit which can rarely be profitable undertakings, but are nevertheless of high value to students in various departments of intellectual activity. This appears to be the aim of the Harvard syndics, as it is of like boards of other universities in the United States.  
From *Nature* 13 February 1913

should explore how, during natural cellular processes, the inhibitory interaction between GAPR-1 and beclin-1 is overcome to induce autophagy.

These data are of therapeutic relevance, particularly with regard to microbial infections and disorders characterized by the accumulation of harmful intracellular proteins. Treatment of cells with Shoji-Kawata and colleagues' pro-autophagic peptide inhibited the replication of several viruses, including human pathogens such as HIV-1, West-Nile virus and chikungunya virus. This antiviral activity was also evident *in vivo* in mouse models of West Nile and chikungunya infections. Moreover, the peptide inhibited the replication of *Listeria monocytogenes* — the intracellular bacterium that causes the infection listeriosis. And it mediated the elimination of harmful protein aggregates associated with the neurodegenerative disorder Huntington's disease.

Thus, the beclin-1-derived peptide could potentially be used for treating diseases in which the induction of autophagy might be beneficial. Nonetheless, before this peptide or any other autophagy inducers can be used in the clinic, it is essential to determine their potential toxicity.

Shoji-Kawata and collaborators observed no significant toxicity in cells treated with therapeutically relevant concentrations of the pro-autophagy peptide, nor do they report any notable side effects of its administration in mice at concentrations that have antiviral activities. However, it remains to be

determined whether the general beneficial effects of increased autophagy for treating infections and other disorders outweigh its possible harmful effects on normal cellular functions. In this respect, strategies that would induce autophagy specifically in infected or damaged cells would be preferable.

Such caveats notwithstanding, this paper clearly illustrates that a better understanding of the mechanisms by which pathogens counteract innate immune responses will make it possible not only to gain insight into the regulation of antimicrobial pathways, but also to devise broad-spectrum therapeutic strategies that potentiate innate immunity. Structural clues obtained about the synthesis of autophagosomes<sup>7</sup> should aid this understanding by providing a more detailed view of the biochemical processes that trigger autophagy. ■

**Adolfo García-Sastre** is at the Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.  
e-mail: [adolfo.garcia-sastre@mssm.edu](mailto:adolfo.garcia-sastre@mssm.edu)

1. Shoji-Kawata, S. *et al.* *Nature* **494**, 201–206 (2013).
2. Fensterl, V. & Sen, G. C. *BioFactors* **35**, 14–20 (2009).
3. Kyei, G. B. *et al.* *J. Cell Biol.* **186**, 255–268 (2009).
4. Gannagé, M. *et al.* *Cell Host Microbe* **6**, 367–380 (2009).
5. Orvedahl, A. *et al.* *Cell Host Microbe* **1**, 23–35 (2007).
6. Lee, J.-S. *et al.* *Nature Cell Biol.* **11**, 1355–1362 (2009).
7. Ragusa, M. J., Stanley, R. E. & Hurley, J. H. *Cell* **151**, 1501–1512 (2012).

### EARTH SCIENCE

## All rise for the case of the missing magma

**A detailed geological analysis of a ridge in the Indian Ocean suggests that compositional variations in Earth's mantle have a surprisingly crucial role in the uplift of a bathymetric bulge along the ridge. SEE ARTICLE P.195**

**JOHN MACLENNAN**

Scientists have long been fascinated by the shape of the sea floor; many are curious to know what the scars and swells on Earth's solid outer skin reveal about our planet's deep interior. The mid-ocean ridges, where tectonic plates pull apart, are topped by a chain of volcanoes that circles the globe. The prevailing view among Earth scientists is that variation in the depth of these ridges is controlled by the temperature of the underlying mantle<sup>1–5</sup>. In places where the ridge rises into shallow water, as it does around Iceland, the mantle is thought to be unusually hot.

On page 195 of this issue, however, Zhou and Dick<sup>6</sup> challenge this model of a purely thermal origin for such variation. Instead, they argue that these bulges along the mid-ocean ridges relate to compositional gradients in the underlying mantle, which are perhaps a relic of gigantic melting events that occurred during the break-up of supercontinents. If the authors' interpretation of their geological observations from a ridge in the Indian Ocean is correct, then their work has important implications for our understanding of the mantle\*.

\*This article and the paper under discussion<sup>6</sup> were published online on 6 February 2013.

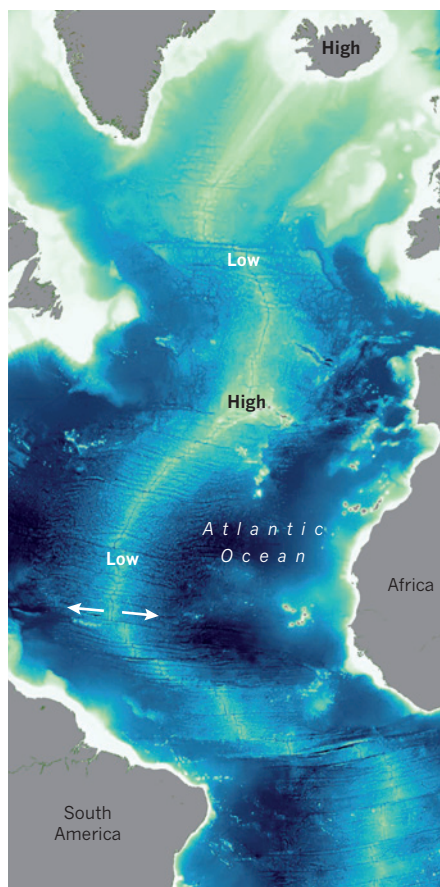
'Armchair' geophysics is now a comfortable reality: the oceans can be surfed on a laptop. Using Google Earth you can see the Mid-Atlantic Ridge, which runs as a long, jagged scar from north to south, roughly following the margins of the continents it separates (Fig. 1). Mantle rock rises up and melts as it decompresses in response to the slow separation of the plates at this and other spreading ridges. The molten rock rises towards the surface, fuelling volcanic eruptions on the ridges and solidifying to form new oceanic crust. The importance of this process is huge: more than two-thirds of Earth's solid surface was created in this way.

If you fix your gaze on a map of the Atlantic at a point midway between France and Newfoundland, you will see that water over the ridge is deep (3 kilometres or more). If you follow the ridge about 1,000 km northwards, you will see the plate boundary gradually shoaling until it reaches Iceland, where it rises above the waves. Such variation in ridge depth, with top-to-bottom amplitudes of several kilometres over stretches of thousands of kilometres, is not restricted to the Atlantic. Notable examples are also found near the Galapagos Islands in the Pacific Ocean and on the ridge between Madagascar and Antarctica in the Indian Ocean. This latter ridge, named the Marion Rise after the French explorer Marc-Joseph Marion du Fresne, is the focus of the study by Zhou and Dick<sup>6</sup>.

Density variations in the mantle are required to support bathymetric bulges such as the Marion Rise. The idea that these variations are thermal in origin gained predominance in the late 1980s, when global surveys found striking correlations between the depths of mid-ocean ridges, the thickness of the oceanic crust and the chemistry of ridge volcanoes<sup>4,5</sup>. The sense of the observed correlation between the depth of mid-ocean ridges and the oceanic-crust thickness — with thick crust associated with shallow ridges — matches the prediction of models in which the mantle that rises under ridges is more or less uniform in composition but varies in temperature by about 200 °C over distances of thousands of kilometres.

High temperatures, perhaps of more than 1,500 °C, are associated with low-density mantle and thick crust because the extent of mantle melting is controlled by temperature and the magma produced by the melting rises to form the crust. The success and simplicity of these models has convinced most Earth scientists of the significance of variations in the mantle temperature. Indeed, a study of the Marion Rise by Cannat *et al.*<sup>7</sup> found a relationship between lava chemistry and ridge depth that is remarkably consistent with the predictions of these models.

The central argument of Zhou and Dick is that temperature variations are not the sole cause of uplift of the Marion Rise. Their detailed survey found widespread exposure



**Figure 1 | Bathymetry of the Atlantic Ocean floor.** The white arrows indicate the separation of the tectonic plates on either side of the Mid-Atlantic Ridge. The depth of the ridge beneath sea level varies substantially, with elevated regions (high) found close to the Azores and Iceland, and deep stretches in-between (low). Although these variations in ridge depth are usually explained in terms of variations in the temperature of the underlying mantle, Zhou and Dick<sup>6</sup> propose that a similar high in the Indian Ocean results from compositional anomalies in Earth's interior. (Map prepared with Generic Mapping Tools<sup>9</sup>, using ETOPO2v2 bathymetric data<sup>10</sup> and position of the Mid-Atlantic Ridge from GPlates software<sup>11</sup>.)

of mantle rocks on the ridge, meaning that the crust is absent or very thin. This missing magma raises questions about a thermal origin for the Marion Rise, which should be associated with 3–6 km of crust on the ridge<sup>7</sup>. Bolstered by this inference of negligible crust, Zhou and Dick conclude that differences in mantle composition keep the Marion Rise elevated above the surrounding stretches of ridge. This support mechanism, whereby past melting events generate a solid residue that has a lower density than typical mantle<sup>8</sup>, should produce an association between shallow ridges and thin crust. Although this mechanism was first proposed<sup>8</sup> in the early 1970s, it later fell out of favour when it failed to account for the observed global-survey relationships between crustal thickness and ridge depths<sup>4,5</sup>. Intriguingly, Zhou and Dick speculate that the chemistry of the unusually

light mantle supporting the Marion Rise may be a solid relic of an enormous melting event that took place just before the break-up of the super-continent Gondwana, 180 million years ago.

One of the great scientific challenges to our understanding of the submerged ocean ridges is the difficulty in producing geological maps of these active zones. Therefore, the use by Zhou and Dick of dense sampling to demonstrate that a single segment on the Marion Rise is more or less missing its crust is a significant contribution. This segment, however, has a water depth of close to 4 km, more than 1.5 km deeper than the shallowest segments of the Marion Rise<sup>7</sup>. An obvious next step in testing their hypothesis would be to survey the rocks of these shallowest parts of the rise. Another source of ambiguity in our understanding of the origin of the Marion Rise is a lack of knowledge of the variations in crustal thickness along it. Unfortunately, the interpretation of seismic data from regions where mantle is pervasively altered by circulation of sea water poses a major challenge. An ambitious and expensive combination of geophysical investigation and deep drilling of the ocean floor would, however, provide clearer observational data on the true variations in crustal thickness along the Marion Rise. With such data in hand, it will then become possible to test Zhou and Dick's model, and to establish whether thermal or compositional variation in the mantle is the predominant factor in controlling the depths of the mid-ocean ridges.

An intriguing aspect of the authors' work is that the inferred anomaly in the mantle composition is thought to be the result of an ancient melting event that occurred when an unusually hot plume of mantle rose under Gondwana. An old thermal anomaly is therefore suggested to have produced a modern compositional structure. Disentangling the signs of ancient events and modern processes remains an imposing challenge for scientists who wish to understand Earth's interior. ■

**John MacLennan** is in the Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK.  
e-mail: jcm1004@cam.ac.uk

1. Wilson, J. T. *Phil. Trans. R. Soc. Lond. A* **258**, 145–167 (1965).
2. Morgan, W. J. *Nature* **230**, 42–43 (1971).
3. Schilling, J.-G. *Nature* **242**, 565–571 (1973).
4. Klein, E. M. & Langmuir, C. H. *J. Geophys. Res.* **92**, 8089–8115 (1987).
5. McKenzie, D. & Bickle, M. J. *J. Petrol.* **29**, 625–679 (1988).
6. Zhou, H. & Dick, H. J. B. *Nature* **494**, 195–200 (2013).
7. Cannat, M. *et al. Geochim. Geophys. Geosyst.* **9**, Q04002 (2008).
8. O'Hara, M. J. *Nature* **253**, 708–710 (1975).
9. Wessel, P. & Smith, W. H. F. *Eos* **79**, 579 (1998).
10. ETOPO2v2 Global Gridded 2-minute Database [www.ngdc.noaa.gov/mgg/global/etopo2.html](http://www.ngdc.noaa.gov/mgg/global/etopo2.html).
11. Williams, S., Müller, R. D., Landgrebe, T. C. W. & Whittaker, J. M. *GSA Today* **22**, <http://dx.doi.org/10.1130/GSATG139A.1> (2012).



# Molecular signatures of G-protein-coupled receptors

A. J. Venkatakrishnan<sup>1</sup>, Xavier Deupi<sup>2</sup>, Guillaume Lebon<sup>1,3,4,5</sup>, Christopher G. Tate<sup>1</sup>, Gebhard F. Schertler<sup>2,6</sup> & M. Madan Babu<sup>1</sup>

**G-protein-coupled receptors (GPCRs) are physiologically important membrane proteins that sense signalling molecules such as hormones and neurotransmitters, and are the targets of several prescribed drugs. Recent exciting developments are providing unprecedented insights into the structure and function of several medically important GPCRs. Here, through a systematic analysis of high-resolution GPCR structures, we uncover a conserved network of non-covalent contacts that defines the GPCR fold. Furthermore, our comparative analysis reveals characteristic features of ligand binding and conformational changes during receptor activation. A holistic understanding that integrates molecular and systems biology of GPCRs holds promise for new therapeutics and personalized medicine.**

Signal transduction is a fundamental biological process that is required to maintain cellular homeostasis and to ensure coordinated cellular activity in all organisms. Membrane proteins at the cell surface serve as the communication interface between the cell's external and internal environments. One of the largest and most diverse membrane protein families is the GPCRs, which are encoded by more than 800 genes in the human genome<sup>1</sup>. GPCRs function by detecting a wide spectrum of extracellular signals, including photons, ions, small organic molecules and entire proteins. After ligand binding, GPCRs undergo conformational changes, causing the activation of complex cytosolic signalling networks, resulting in a cellular response. Altering the activities of GPCRs through drugs is already used in the treatment of numerous ailments including cardiac malfunction, asthma and migraines. Given the tremendous diversity of GPCRs, there remains enormous potential for the development of additional drugs to ameliorate neurological disorders, inflammatory diseases, cancer and metabolic imbalances. Thus, determining the structure of GPCRs and understanding the molecular mechanism of receptor activation is not only of fundamental biological interest, but also holds great potential for enhancing human health.

In accordance with the guidelines of the International Union of Basic and Clinical Pharmacology, non-sensory GPCRs (that is, excluding light, odour and taste receptors) can be classified, according to their pharmacological properties, into four main families: class A rhodopsin-like, class B secretin-like, class C metabotropic glutamate/pheromone, and frizzled receptors. In the past 12 years, more than 75 structures of 18 different class A GPCRs have been determined in complex with ligands of varied pharmacology, peptides, antibodies and a G protein (Fig. 1). These structures provide unprecedented insights into the structural and functional diversity of this protein family. Given the recent exciting advances in the field of GPCR structural biology, we are in a unique position to address the following fundamental questions: what are the 'molecular signatures' of the GPCR fold? And what are the molecular changes that the receptor undergoes during activation? The availability of structures of diverse GPCRs now permits a systematic comparative analysis of the GPCR fold. This knowledge, combined with the understanding gained from complementary biophysical, computational, and biochemical studies empowers us to probe the molecular basis of GPCR structure–function relationship

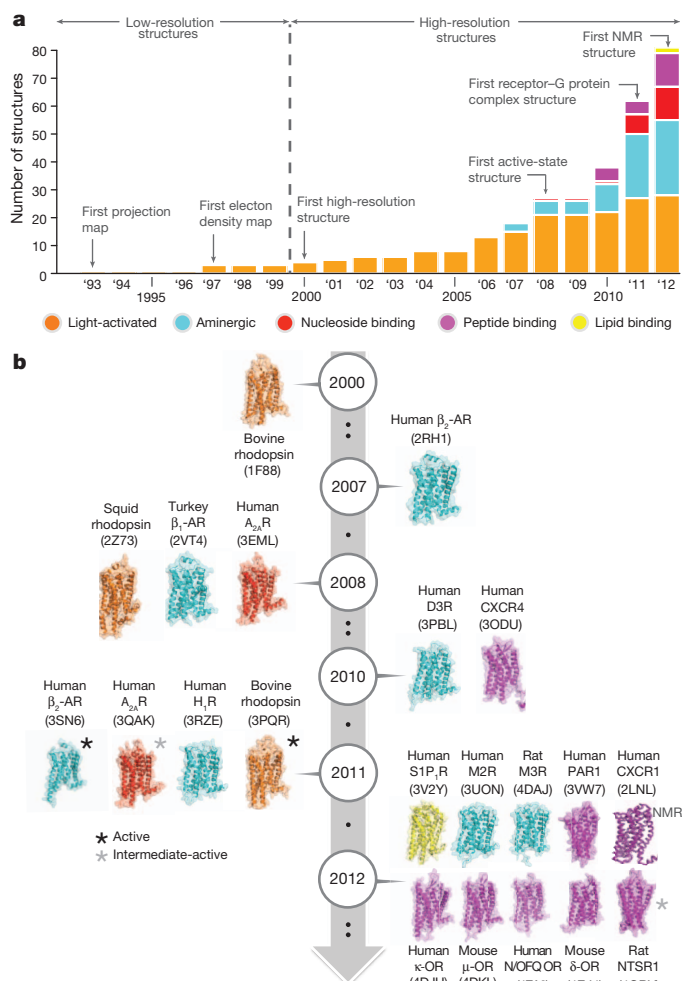
comprehensively, and in the process expand the current frontiers of GPCR biology.

In this analysis, we objectively compare known structures and reveal key similarities and differences among diverse GPCRs. We identify a consensus structural scaffold of GPCRs that is constituted by a network of non-covalent contacts between residues on the transmembrane (TM) helices. By systematically analysing structures of the different receptor–ligand complexes, we identify a consensus 'ligand-binding cradle' that constitutes the bottom of the ligand-binding pocket within the TM bundle. Furthermore, our comparative study suggests that the third TM helix (TM3) has a central role as a structural and functional hub. We then synthesize our current understanding of the structural changes during activation by discussing conformational dynamics of the receptor as gleaned from recent biochemical, biophysical and computational studies. Finally, we highlight open challenges in the field and discuss exciting new directions for GPCR research.

## Reasons for GPCR crystallography successes

Since 2007, several innovative protein engineering techniques and crystallography methods<sup>2</sup> have resulted in an almost exponential growth in the number of solved structures (Fig. 1a). These include creating receptor–T4 lysozyme<sup>3,4</sup> and receptor–apocytochrome<sup>5</sup> chimaeras, co-crystallization with monoclonal antibody fragments from either mouse<sup>6,7</sup> or camelids<sup>8,9</sup>, and thermostabilization of GPCRs by systematic scanning mutagenesis<sup>10–12</sup> or by engineering disulphide bridges<sup>13,14</sup>. Often, it was necessary to truncate flexible regions of the receptor and to use high-affinity/low off-rate ligands to enhance receptor stability. In addition, the use of lipidic cubic phase<sup>15</sup> and new detergents<sup>16</sup> has improved the likelihood of obtaining crystals, and advances in micro-crystallography have allowed obtaining higher resolution diffraction from smaller crystals<sup>17</sup>. Although these methods have been successful, the inherent limitations with such recombinant methods are that (1) post-translational modifications are often removed or not incorporated during protein purification; (2) truncations of loops and amino or carboxy termini provide limited understanding of the structure and function of these regions; and (3) insertion of, for example, T4 lysozyme into intracellular loop 3, addition of thermostabilizing mutations or the use of antibodies may affect the relative mobility of the TM helices, thereby potentially biasing the conformational ensemble of the receptor. Despite the above-mentioned

<sup>1</sup>MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK. <sup>2</sup>Paul Scherrer Institute, 5232 Villigen PSI, Switzerland. <sup>3</sup>CNRS, UMR-5203, Institut de Génétique Fonctionnelle, Montpellier F-34094, France. <sup>4</sup>INSERM, U661, Montpellier F-34094, France. <sup>5</sup>Universités de Montpellier 1 & 2, UMR-5203, Montpellier F-34094, France. <sup>6</sup>Department of Biology, ETH Zürich, Wolfgang-Pauli-Str. 27, 8093 Zürich, Switzerland.



**Figure 1 | Time-line of GPCR structures.** **a**, Bar chart showing the increase in the number of GPCR structures with time. **b**, Time-line showing representative crystal structures of GPCRs and the year of publication. Active conformations are marked with a black asterisk, and an intermediate-active conformation is marked with a grey asterisk. Protein Data Bank accession numbers are shown in parentheses.

shortcomings, the wide variety of techniques is likely to ensure a steady increase in the number of GPCR structures in the future.

## Solved structures of GPCRs

So far, high-resolution structures have been solved for the following class A GPCRs (Supplementary Table 1): (1) rhodopsin (bovine rhodopsin<sup>18</sup> and squid rhodopsin<sup>19</sup>); (2) several members of aminergic GPCRs:  $\beta$ -adrenoceptors (avian  $\beta_1$ -AR (ref. 10) and human  $\beta_2$ -AR (refs 3, 6)), muscarinic acetylcholine receptors (human M2R (ref. 20) and rat M3R (ref. 21)), human  $H_1$  histamine receptor<sup>22</sup>, and human D3 dopamine receptor<sup>23</sup>; (3) a nucleoside-binding GPCR: human adenosine  $A_{2A}$  receptor ( $A_{2A}$ R)<sup>24</sup>; (4) several members of the peptide-binding GPCRs: human CXCR4 chemokine receptor<sup>25</sup>, opioid receptors (human nociceptin receptor<sup>5</sup> and  $\kappa$ -OR<sup>26</sup> and mouse  $\mu$ -OR<sup>27</sup> and  $\delta$ -OR<sup>28</sup>), rat neurotensin receptor (NTSR1)<sup>29</sup> and human protease-activated receptor (PAR1)<sup>30</sup>; and (5) a lipid-binding GPCR: human sphingosine-1 phosphate (S1P<sub>1</sub>) receptor<sup>31</sup>. The human CXCR1 chemokine receptor is the first GPCR structure that was determined using NMR spectroscopy<sup>32</sup>.

The crystal structures of all of the above-mentioned class A GPCRs (except NTSR1) have been obtained in inactive conformations bound to either inverse agonists that reduce basal activity or neutral antagonists that maintain basal activity. Rat NTSR1 (ref. 29), bovine rhodopsin<sup>13,14,33</sup>, human  $\beta_2$ -AR (refs 8, 9, 34), avian  $\beta_1$ -AR (refs 35, 36) and human  $A_{2A}$ AR

(refs 11, 37) were crystallized with agonists that induce an increase in biological activity. Of these, only bovine rhodopsin<sup>13,14,33,38</sup>, human  $\beta_2$ -AR (refs 8, 9), human  $A_{2A}$ AR (refs 11, 37) and rat NTSR1 (ref. 29) were obtained in active (or intermediate-active) states. An important landmark in GPCR biology was the determination of the active-state ternary complex of  $\beta_2$ -AR in complex with the heterotrimeric G protein<sup>9</sup>.

## Molecular signatures of the GPCR fold

The structure of a GPCR can be divided into three parts: (1) the extracellular region, consisting of the N terminus and three extracellular loops (ECL1–ECL3); (2) the TM region, consisting of seven  $\alpha$ -helices (TM1–TM7); and (3) the intracellular region, consisting of three intracellular loops (ICL1–ICL3), an intracellular amphipathic helix (H8), and the C terminus (Fig. 2a). In a broad sense, the extracellular region modulates ligand access; the TM region forms the structural core, binds ligands and transduces this information to the intracellular region through conformational changes, and the intracellular region interfaces with cytosolic signalling proteins.

## Extracellular region and ligand-binding pocket accessibility

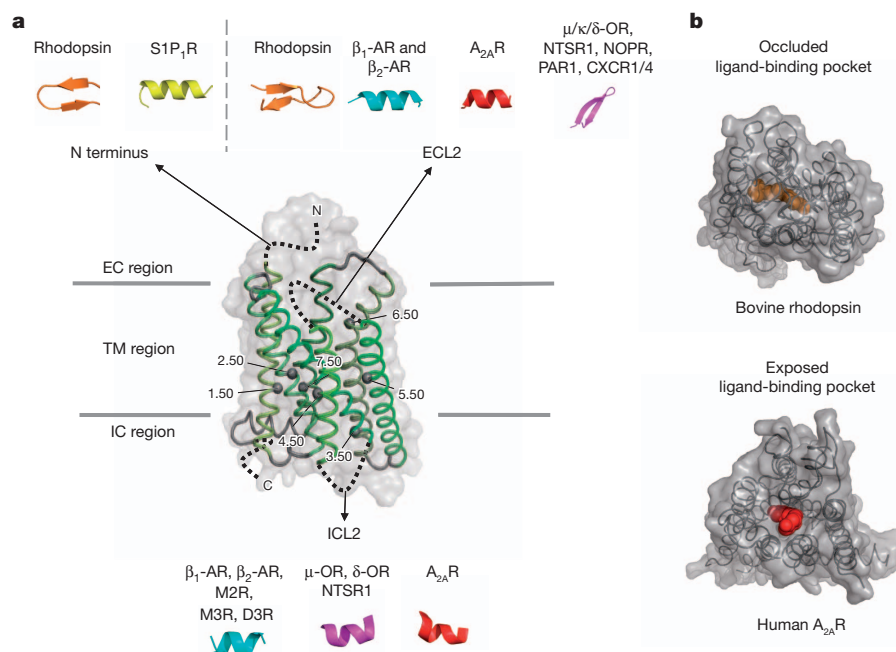
Sequence analysis shows that there is a large diversity in the lengths and sequence compositions of the N terminus<sup>39</sup> and the extracellular loops<sup>40</sup>. The class A GPCR structures reveal two distinct types of extracellular region: those that either occlude the ligand-binding pocket or leave the ligand-binding pocket water-accessible (Fig. 2b). Rhodopsin<sup>18</sup> and the S1P<sub>1</sub> receptor<sup>31</sup> have occluded binding pockets, presumably because they both bind hydrophobic ligands that may enter the receptor from the lipid bilayer<sup>41</sup>. The N terminus and ECL2 of rhodopsin fold into  $\beta$ -hairpin loops, and together they form a 'lid' for the ligand-binding pocket. Similarly, the S1P<sub>1</sub> receptor contains a three-turn  $\alpha$ -helix that packs against ECL2 and ECL3 (ref. 31). In the receptors that bind water-soluble ligands, ECL2 can differ structurally between receptors, but the structures are likely to be conserved in a subfamily-specific manner. ECL2 can contain helices (for example, certain aminergic or adenosine receptors) or sheets (for example, peptide-binding receptors) (Fig. 2a). Even in the absence of defined secondary structural elements in ECL2 (for example, in muscarinic receptors), it still partially folds over the extracellular region and shapes the route for ligand entry into the binding pocket. Indeed, molecular dynamics simulations suggest that ECL2 could be involved in the first steps of ligand recognition and selectivity in the  $\beta$ -ARs<sup>21,42,43</sup>. Furthermore, pharmacological studies have shown that this region is important for ligand-binding kinetics<sup>27</sup>. In contrast to ECL2, ECL1 and ECL3 are relatively short and tend to lack distinct secondary structural elements<sup>40</sup>.

A unique feature of the extracellular region is the presence of disulphide bridges that contribute to receptor stability. Although there are several subfamily-specific disulphide bridges, the one between a crucial residue in TM3, Cys<sup>3.25</sup> (in which the superscript denotes Ballesteros–Weinstein numbering<sup>44</sup>), and ECL2 seems to be highly conserved in most GPCR structures (except S1P<sub>1</sub>). This TM3–ECL2 disulphide bridge anchors the extracellular side of the helix near the binding site, and limits the extent of the conformational changes of this region during receptor activation. Indeed, reducing the disulphide bridges can influence receptor stability and activity. Furthermore, in several GPCRs, ECL3 contains an additional intra-loop disulphide bridge within a CX<sub>n</sub>C motif that possibly influences receptor function by limiting the conformational freedom available to the loop. For instance, a missense mutation (Cys271Arg) in this disulphide bridge in the melanocortin-4 receptor results in receptor malfunction and is linked to obesity.

## Conserved structural scaffold in the TM region

The TM helix bundle serves as the communication link between the ligand-binding pocket and the G-protein-coupling region. Although GPCRs share a similar architecture of seven TM helices held together by tertiary contacts, their sequences are diverse. An objective comparison of the structures of diverse GPCRs using a network representation permits





**Figure 2 | Diversity in the secondary structure elements of GPCRs in the extracellular and intracellular regions.** **a**, TM helices (TM1–TM7) are shown as cartoon (coloured in a spectrum of green) and surface representation. Numbers denote Ballesteros–Weinstein numbering. In this receptor-independent notation, each residue is identified by two numbers that are separated by a dot (for example, 1.50): the first number ranges from 1 to 7 and corresponds to the TM helix where the residue is located; the second number indicates its position relative to the most conserved residue of the helix, which is assigned the number 50. Residues in the same TM helix that are N- or C-terminal to the '50' residue are assigned a number that decreases or increases

sequentially, respectively. For example, a residue just before or after the most highly conserved residue in TM1 will be assigned 1.49 or 1.51, respectively. N and C termini and the segments containing defined secondary structure in the extracellular (EC) and intracellular (IC) region are shown as dotted lines; the type of secondary structure element for the different representative GPCRs are shown in the grey panels. The loop regions lacking an  $\alpha$ -helix or a  $\beta$ -sheet in any of the structures are not shown. See Fig. 1 for receptor colour code.

**b**, Extracellular region that occludes or exposes the ligand-binding pocket as seen from the extracellular side (top view).

us to investigate whether any tertiary contacts between TM helices are conserved, independent of sequence diversity. A systematic analysis of the different GPCR structures, which includes both active and inactive states, reveals a consensus network of 24 inter-TM contacts mediated by 36 topologically equivalent amino acids (Supplementary Table 2). The topologically equivalent positions are identified through structure-based sequence alignment and are referred to by the Ballesteros–Weinstein numbering scheme (Supplementary Table 3). In this consensus network, the contacts are present in all (or all but one) of the structures, irrespective of their conformational state, and thus are likely to represent structurally important positions in the receptor (Fig. 3). The importance of these positions is highlighted by the fact that mutations in 14 out of 36 positions have been noted to result in either an increase or a loss of receptor activity<sup>45</sup>. With the availability of more high-resolution structures of other GPCRs, one may converge on a unified subset of inter-helical contacts that is maintained in all GPCRs.

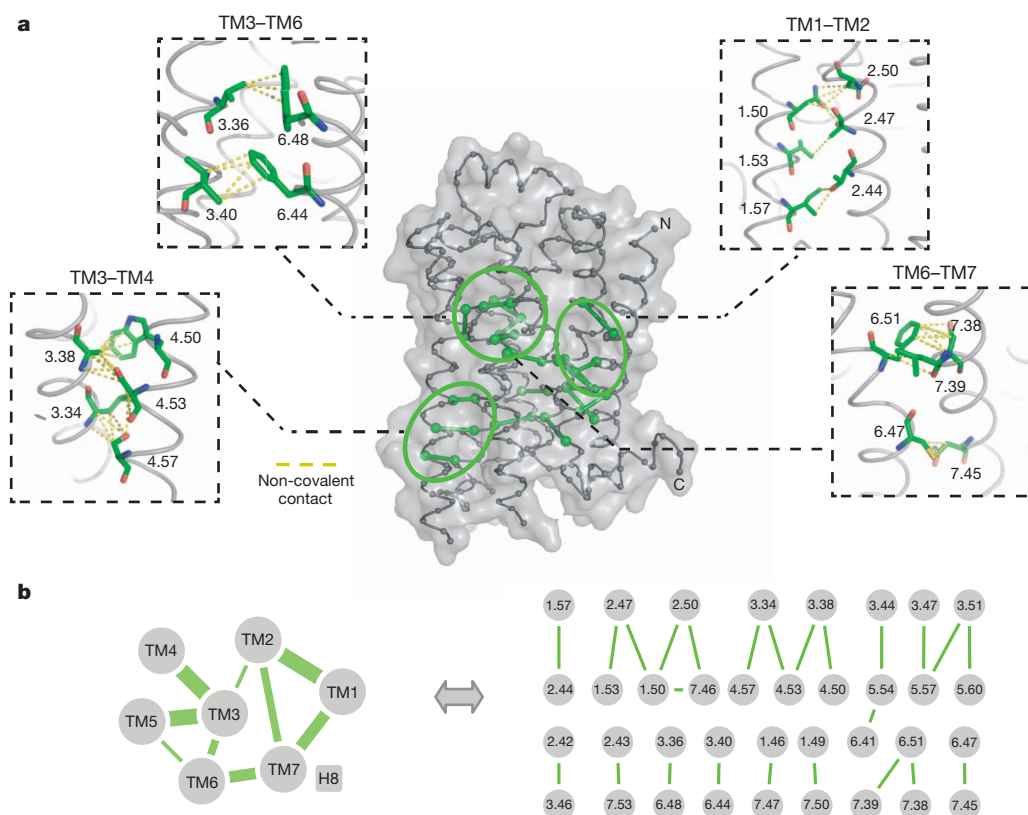
The 36 topologically equivalent residues of the structural scaffold include highly conserved residues such as Asn<sup>1.50</sup>, Asp<sup>2.50</sup>, Trp<sup>4.50</sup> and Pro<sup>7.50</sup>. Nevertheless, we also observe that many topologically equivalent positions can tolerate variability in amino acid substitutions. The identity of some of these 36 positions may be variable, but they all nevertheless predominantly maintain the non-covalent contacts between them. For instance, although a contact between 2.42 and 3.46 is seen in all structures, these residues are different among different receptors: for example, Ile 75 and Leu 131 in bovine rhodopsin and Tyr 97 and Met 152 in the human  $\kappa$ -OR. Thus, the consensus inter-TM contact network seems to provide an evolutionarily conserved structural scaffold of non-covalent contacts for the GPCR fold. It is likely that the tolerance of sequence variability in some of these positions permits diverse sequences to adopt a similar structure, thereby contributing to the evolutionary success of the GPCR fold. These conformation-independent consensus contacts may constitute a rigid platform on which distinct conformation-specific

structural changes take place. Importantly, the network approach used here and the consensus set of inter-TM tertiary contacts identified should be valuable for GPCR engineering, *de novo* GPCR modelling and to increase the accuracy of GPCR homology models for various applications<sup>46</sup>.

In terms of spatial positioning within the receptor, the consensus inter-TM tertiary contacts are largely localized to the central and cytoplasmic side of the TM bundle and primarily clustered at the interfaces of TM1–TM2, TM3–TM4, TM3–TM5 and TM3–TM6–TM7. Conservation of these contacts across diverse GPCRs may be due to the requirement for receptor biogenesis, protein stability or functionality. For instance, TM1 and TM2 do not undergo any major movement after receptor activation. Because they are the first two TM regions to be translated by the ribosome, the consensus contacts observed here might have an important role in membrane insertion, folding and topogenesis of GPCRs. Indeed, mutagenesis experiments of the neurotensin receptor<sup>47</sup> provide support for this possibility. Similarly, TM3 shares consensus helical packing interfaces with all other TM helices except TM1 and TM7, suggesting a role in maintaining the fold (Fig. 3b). This does not mean that TM3 makes no contacts with TM1 or TM7 but that the contacts between equivalent residues are not maintained across different receptors. Whereas the middle portion of TM3 makes consensus contacts with TM4 and TM6, the portion towards the cytoplasm makes contacts with TM5 and TM2. Thus, TM3 seems to have a key role of 'structural hub' in maintaining the scaffold in all GPCR structures, both in the inactive and active conformational states (Fig. 3b). Intriguingly, TM3 adopts an extreme tilt-angle ( $\sim 35^\circ$ ), and this unusual geometry may facilitate its role as a structural hub.

### Consensus scaffold of class A GPCR ligand-binding pocket

A remarkable feature of the GPCR family is its ability to bind ligands of diverse shapes, sizes and chemical properties. Although all ligands have



**Figure 3 | Consensus scaffold of non-covalent contacts in GPCRs.**

**a**, Network of 24 inter-helical contacts between 36 topologically equivalent residues is shown on a representative structure of inactive  $\beta_1$ -AR. The spatially clustered contacts between the amino acids are shown in the panels; to maintain visual clarity, contacts between TM3 and TM5 are not shown. Here we define that a pair of residues is in contact if the Euclidean distance between any pair of atoms (side-chain and/or main-chain atoms) is within the van der Waal interaction distance (that is, the sum of the van der Waal radii of the atoms plus 0.6 Å). With the availability of more high-resolution structures of other GPCRs,

been observed to bind in a pocket in the extracellular side of the TM bundle, different ligands penetrate to different depths within this pocket (Fig. 4a). Despite the diversity in the ligands, a systematic comparison of the residues that contact the ligand revealed similarities in the ligand-binding pocket (that is, residues within 4 Å distance of any ligand atom). We observe that except for the ligands of the CXCR4 receptor and NTSR1, most of the ligand-contacting residues are present in the TM helices (Fig. 4b). Topologically equivalent residues from TM3, TM6 and TM7 typically contact the ligand in nearly all receptors. In particular, residues at positions 3.32, 3.33, 3.36, 6.48, 6.51 and 7.39 make consensus contacts with diverse ligands across class A GPCRs. Residues from the other TMs tend to contact specific ligands to different extents. Thus, one can surmise that these key positions in TM3, TM6 and TM7 form a consensus scaffold of the ligand-binding pocket and that variation in the amino acids occupying the topologically equivalent positions contribute to ligand specificity in different receptors. TM1 is not directly involved in contacting the ligand in any structure, suggesting that mutations in TM1 that affect ligand binding are likely to be indirect. In addition to residues that directly contact ligand, water molecules have also been observed to mediate indirect contacts between the ligand and the receptor<sup>11,24,37,48</sup>.

#### The ligand-binding pocket and consensus scaffold interface

Of the positions that form the consensus ligand-binding pocket, two pairs of contacting residues (between 3.36–6.48 and 6.51–7.39) are also present in the consensus inter-TM contact network (Fig. 4b). Indeed, biochemical studies have shown that mutations in these positions affect

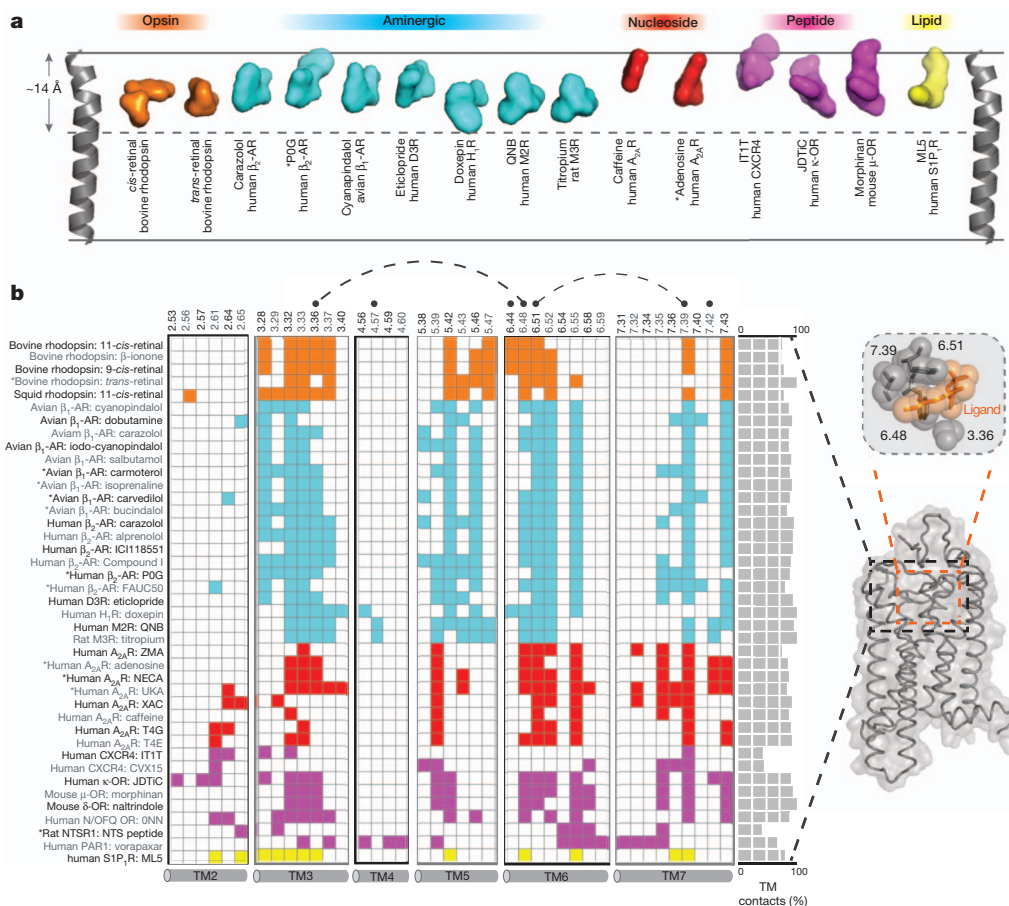
one may converge on a unified subset of inter-helical contacts that is maintained in all GPCRs. **b**, Schematic representation of the inter-TM tertiary contact network in which TM helices are presented as circles, and H8 is shown as a square. The lines between a pair of circles indicate the presence of consensus contacts, and the thickness of the line is proportional to the number of contacts between the TM helices. In the right panel, the thickness of lines is kept uniform because here a line indicates only the presence of a non-covalent contact and not the number of non-covalent contacts. Numbers denote Ballesteros–Weinstein numbering (see Fig. 2 legend).

receptor conformational selectivity<sup>12</sup> and ligand-binding affinity<sup>49</sup>. Furthermore, a non-covalent contact between residues 6.48 and 6.51 in TM6 is also maintained in the structures. Together, residues in these positions appear to form a ‘ligand-binding cradle’ in the TM bundle in nearly all class A GPCR structures. An important role of the conserved structural scaffold might be to position the ligand-binding region of the TM helices in a precise configuration that forms a pocket. Importantly, this set of consensus ligand-binding positions can further our understanding of the ligand-binding pockets of other GPCRs, the molecular basis of cross-reactivity, the specificity of ligands among subfamilies of receptors, poly-pharmacology and also aid in fragment-based drug discovery.

#### Functional and structural importance of intracellular regions

Residues in the intracellular region and the cytoplasmic ends of TM regions bind downstream signalling effectors such as G proteins, GPCR kinases and arrestins<sup>9,50</sup>. In the available structures, ICL1 is typically six amino acids long and contains a helical turn, whereas ICL2 generally has either a one- or two-turn  $\alpha$ -helix or an unstructured stretch of amino acids (Fig. 2). Interestingly, in all the opioid receptor and muscarinic receptor structures, a conserved Arg in ICL2 forms a salt bridge with Asp<sup>3.49</sup> of the DRY motif in TM3, thereby tethering ICL2 with the TM core<sup>5</sup>. This is analogous to a conserved Tyr in ICL2 of several aminergic receptors that also forms a hydrogen bond with Asp<sup>3.49</sup>. The importance of this interaction is highlighted by the fact that mutation of Tyr 149 in the ICL2 of avian  $\beta_1$ -AR decreases receptor stability<sup>10</sup>, and phosphorylation of Tyr 141 in the ICL2 of human  $\beta_2$ -AR





**Figure 4 | Ligand-binding pocket in class A GPCRs.** **a**, Comparison of depth of ligand-binding pockets in which TM4 is used as a frame of reference. The depth of ligand penetration into the TM bundle is the deepest for doxepin in the histamine H<sub>1</sub> receptor and shallowest for caffeine in A<sub>2A</sub>AR. The label on the top indicates the subfamily of class A GPCRs. **b**, Characterization of ligand-binding pockets of class A GPCRs. Comparison of the TM residues that are present in the ligand-binding pocket is shown as a matrix. Receptor–ligand information is shown as rows, and the Ballesteros–Weinstein numbers of TM residues that contact the ligand are shown as columns. The TM residues that are present in the consensus inter-TM contact network are marked with black dots, and

contacts between these residues are shown as dotted lines. Rows marked with an asterisk denote agonist-bound receptor structures. In the matrix, the presence of a contact between the ligand and the TM residue is shown as a coloured box, and the absence of a contact is shown as an empty box. The percentage of TM residue contacts made by the ligand is shown as a bar plot on the right of the matrix (other contacts that are not shown are made with residues in the extracellular region). The backbone representation of rhodopsin is shown as a reference for showing the ligand-binding pocket (black dashed square). The consensus ligand-binding pocket, also referred to as the ligand-binding cradle, is illustrated with retinal as an example (orange dashed box).

facilitates a shift of the receptor conformational equilibrium towards the active state<sup>51</sup>. In the  $\beta_2$ -AR–G-protein complex, ICL2 was observed to interact with the N terminus of the G $\alpha$  subunit<sup>9</sup>. In addition to the ICLs, a short amphipathic helix (H8), typically three turns long and with palmitoylation sites at its C terminus, is present in several class A GPCR structures with the exception of the CXCR4, NTSR1 and PAR1, in which this region was observed to be unstructured. Although several studies have implicated H8 in G-protein binding, this region is not seen to contact the G protein in the structure of the GPCR heterotrimeric G-protein ternary complex<sup>9</sup>. Future studies on H8 should provide insights into its role in the structure and function of the receptor.

### Intrinsically disordered segments in intracellular regions

ICL3 and the C-terminal tail are long, and variable regions<sup>40</sup> and are probably intrinsically disordered in many GPCRs<sup>52</sup>. Such disordered regions typically expose linear peptide motifs that recognize specific partners and allow for regulation of their binding and function<sup>53,54</sup>. Interestingly, a polybasic motif proximal to H8 in many GPCRs has been shown to facilitate G-protein pre-coupling and influence the rate of receptor activation<sup>55</sup>. Furthermore, several residues in the C-terminal tail of  $\beta_2$ -AR (ref. 56) are extensively post-translationally modified, providing support for the existence of a signalling 'bar code'. In

$\beta_2$ -AR, diverse GPCR receptor kinases and insulin receptor tyrosine kinase can phosphorylate the cytoplasmic regions, and the different phosphorylated forms of the receptor show distinct patterns of interaction with  $\beta$ -arrestin, thereby influencing receptor activity and internalization from the membrane<sup>56</sup>. A more thorough investigation of the residues in the disordered regions may provide important insights into the diversity of downstream interaction partners and signalling events.

### TM3 is a structural and functional hub

Considering the observations from comparing the structures of diverse GPCRs in the active and inactive states, it emerges that almost every position in TM3 seems to be important for maintaining either the structure or function of GPCRs. Such a role may be facilitated by the unusually large tilt-angle of TM3 with respect to the axis perpendicular to the plane of the lipid bilayer. The residues in TM3 form a consensus network of conformation-independent inter-TM contacts between strategically placed residues in other TM helices, thereby defining the GPCR fold (Fig. 3). The extracellular end of TM3 forms a conserved disulphide bridge with ECL2, and Asp<sup>3.49</sup> in the cytoplasmic end interacts with ICL2 in several receptors. At the same time, residues in TM3 mediate extensive contacts with ligands of diverse pharmacology and participate in the formation of the 'ligand-binding cradle' (Fig. 4). After

activation, the cytoplasmic end of TM3 forms an important interface for G-protein binding (Fig. 5) and the Arg<sup>3.50</sup> of the DRY motif directly interacts with a backbone carbonyl of the C terminus of the G protein<sup>9,14</sup>. Importantly, mutations in many of the positions in TM3 cause receptor inactivation or constitutive activation, suggesting that TM3 is a structural and functional hub in GPCRs (Fig. 6).

### Molecular changes during receptor activation

Receptor activation involves binding of ligands to the extracellular part of the TM region and the extracellular region, thereby resulting in small conformational changes in the TM core. This ultimately leads to larger structural rearrangements in the cytoplasmic side of the transmembrane–intracellular interface, facilitating the binding of intracellular effectors to the intracellular region. The active state of a GPCR is thus defined as the conformation of the receptor that couples to and stabilizes an effector molecule such as the heterotrimeric G protein<sup>57</sup>. The availability of structures in the intermediate-active and active states has provided important insights into this general mechanism<sup>29,58–60</sup>. However, this ‘static’ information from crystal structures provides only partial insights into dynamic aspects of activation such as allostery. Several recent studies that interpret biochemical, biophysical and computational techniques in light of the crystal structures are beginning to provide detailed insights into these dynamic processes and the complex equilibrium between the receptor conformational ensembles.

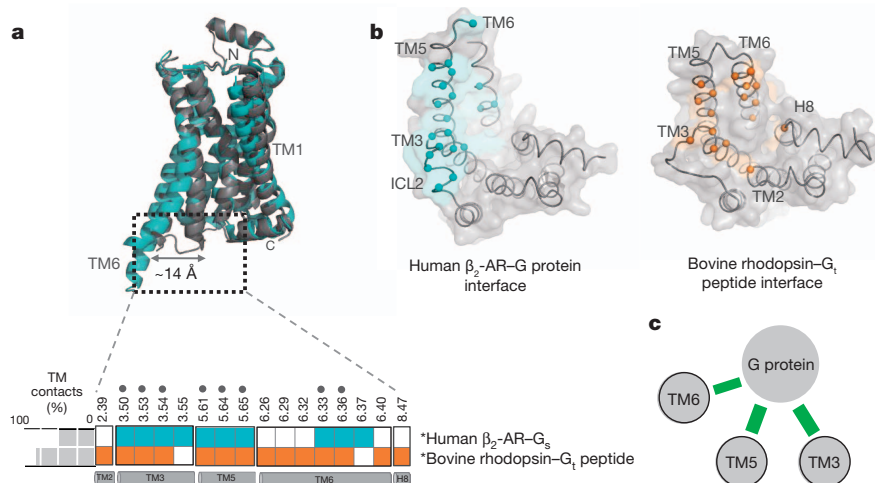
### Existence of several conformational states

Our understanding of conformational dynamics and the activation pathway in GPCRs is primarily derived from experiments carried out on bovine rhodopsin and human  $\beta_2$ -AR using various biophysical and biochemical approaches. In rhodopsin, a series of spectroscopically distinct intermediates has been structurally characterized (Supplementary Table 1). For  $\beta_2$ -AR, it has been shown that binding of agonists stabilizes specific sub-states that are typically sparsely populated in the ligand-free receptor<sup>61</sup>, and also that ligands with different efficacies are able to shift the equilibrium between receptor conformations to different extents<sup>62</sup>. The existence of an intermediate receptor conformation is also supported by atomistic molecular dynamics simulations of  $\beta_2$ -AR (ref. 63). After stabilization of these lowly populated sub-states by ligand

binding, activation proceeds through a series of discrete conformational intermediates<sup>58–60</sup> that ultimately lead to larger structural rearrangements near the transmembrane–intracellular interface.

Conformational changes can be investigated using quantitative mass spectrometry that measures the reactivity of side chains of individual amino acids to particular chemicals. In  $\beta_2$ -AR, it was shown that when receptors bind functionally similar ligands, there were distinct patterns of reactivity of the probe with different amino acids<sup>64</sup>. This supports the view that after binding, a considerable variability in receptor conformation is induced or stabilized by similar ligands. Concordantly, when  $\beta_2$ -AR was studied by hydrogen-deuterium exchange mass spectrometry, it was shown that inverse agonists are more stabilizing whereas agonists induce the largest degree of conformational mobility<sup>65</sup>. Although these findings are consistent with the X-ray structures of the active states, some of the structures of  $\beta$ -ARs bound to agonists were observed to be in the inactive conformation<sup>34–36</sup>. This suggests that agonist binding alone may not be sufficient to stabilize fully active states (as defined by the structure of the  $\beta_2$ -AR–G<sub>s</sub> complex<sup>9</sup>), and the conformation adopted by an agonist-bound receptor depends on the energy landscape for each specific receptor. Only after binding of a G protein, arrestin or conformation-specific antibodies will the fully active state of the receptor become the dominant state. An exception is that it was possible to obtain a ligand-free structure of opsin with an activated conformation in the presence of lipids at low pH<sup>66</sup>.

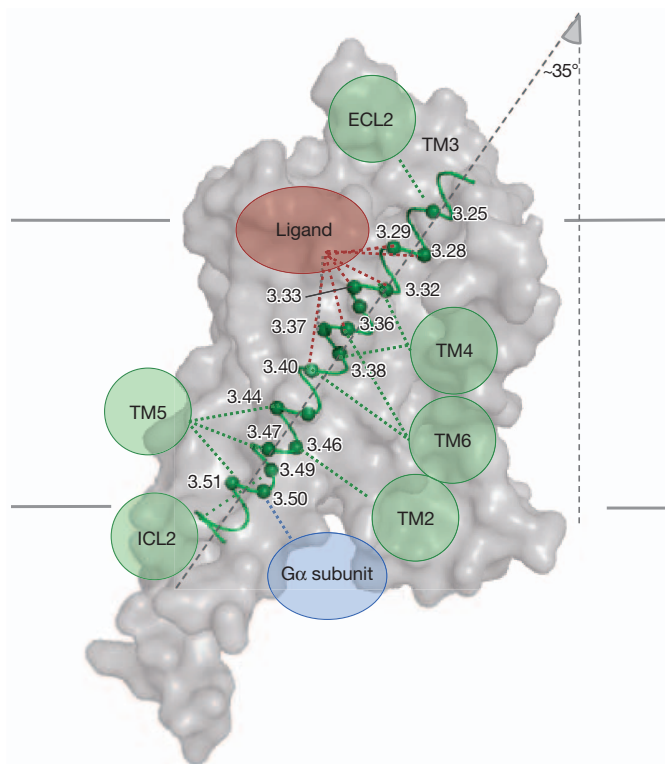
Sequence analysis of GPCR family members has identified a network of co-evolving residues that maybe important for allosteric communication<sup>67</sup>. Furthermore, molecular dynamic simulations of  $\beta_2$ -AR also suggest that a loosely coupled allosteric network links small perturbations at the ligand-binding site to large conformational changes at the intracellular G-protein-binding site, and provide insights into metastable states that may be difficult to study experimentally<sup>63</sup>. Along these lines, computational studies have shown that ligands with different efficacies modulate the free-energy landscape of the receptor by shifting the conformational equilibrium towards active or inactive conformations depending on their elicited physiological response<sup>68,69</sup>. Thus it seems that there may not be a single active state and that ligands can stabilize distinct conformations, thereby giving rise to diverse downstream responses.



**Figure 5 | Characterization of G-protein-binding region within TM helices of GPCRs.** **a**, Comparison of the TM movement between the active (cyan) and inactive (grey) state of  $\beta_2$ -AR (top). TM6 moves 14 Å after activation. TM residues that contact the G<sub>s</sub> protein and G<sub>t</sub> peptide are shown as a matrix for  $\beta_2$ -AR (cyan) and rhodopsin (orange), respectively (bottom). Black dots above Ballesteros–Weinstein numbers indicate residues in the TM region that contact cytoplasmic-binding partners (G<sub>s</sub> protein and G<sub>t</sub> peptide) in both  $\beta_2$ -AR (cyan) and rhodopsin (orange). The percentage of all G<sub>s</sub> protein or G<sub>t</sub> peptide contacting residues within the TM region is shown as a bar plot to the left of the

matrix. Other contacts (not shown) are made with residues in the intracellular region; for instance, ICL2 makes contacts with the N-terminal region of the G $\alpha$  subunit<sup>9</sup>. **b**, Cytoplasmic view of  $\beta_2$ -AR and rhodopsin showing the interface of the G-protein interaction. C $\alpha$  positions of residues in the receptor are shown as spheres. **c**, Schematic representation of the interaction of TM helices with the G protein (contacts with intracellular region are not shown). G protein and TM helices are shown as circles and the interactions are shown as green lines. The thickness of the line is proportional to the number of consensus contacts that are made at the interface of  $\beta_2$ -AR–G<sub>s</sub> protein and rhodopsin–G<sub>t</sub> peptide.





**Figure 6 | Structural and functional hub role of TM3.** The backbone of TM3 is shown as a ribbon helix in dark green. Ballesteros-Weinstein numbers of residues involved in making contacts with various parts of the receptor, ligand and the G protein are shown. Other TM helices and extracellular and intracellular regions interacting with TM3 are shown as green circles. Ligand and G protein are shown as a brown and blue ellipse, respectively. Note the unusual tilt-angle of TM3 with respect to the lipid bilayer.

### Changes in the EC region during receptor activation

The EC region is unlikely to be a passive segment as NMR experiments have revealed conformational changes during receptor activation<sup>70</sup>. In rhodopsin, solid-state NMR data show that activation, which is initiated by the light-induced isomerization of 11-*cis*-retinal to all-*trans*-retinal, is accompanied by conformational changes in ECL2 (ref. 71). In  $\beta_2$ -AR, the existence of different conformations of a salt bridge connecting ECL2 and ECL3 lining the ligand-binding pocket, for agonists, neutral antagonists and inverse agonists was shown based on solution NMR spectroscopy using a modified lysine residue as a conformational probe<sup>72</sup>. These studies suggest that drugs targeting this surface could function as allosteric modulators with GPCR subtype-specific selectivity<sup>72</sup>. In this context, it has been reported that many of the autoantibodies that behave like agonists recognize the extracellular region of the receptor in diverse human diseases such as in Graves' disease. One possibility is that these antibodies stabilize distinct states of the extracellular loop, thereby allosterically modulating receptor activity, leading to activation of downstream signalling pathways in the absence of the ligand<sup>40</sup>.

### TM changes during receptor activation after agonist binding

The agonist-bound structures of rhodopsin<sup>13,14,33</sup>,  $\beta_2$ -AR (refs 9, 34) and  $A_{2A}$ AR (refs 11, 37) show different patterns of ligand-receptor interactions. However, these interactions seem to result in a set of common structural rearrangements in the extracellular part of the TM bundle near the agonist-binding site. These changes can be summarized as follows: (1) small local structural changes in the Pro<sup>5.50</sup>-induced distortion of TM5; (2) relocation of TM3 and TM7; and (3) translation/rotation of TM5 and TM6. In  $\beta_2$ -AR and  $A_{2A}$ AR, agonist binding 'pulls' the extracellular sides of TM3, TM5 and TM7 together, whereas in rhodopsin retinal isomerization

results in an increase in the volume of the binding pocket<sup>14</sup>. Furthermore, in  $A_{2A}$ AR, agonist binding induces a 2 Å shift of TM3 along its axis and towards the extracellular side<sup>11,73</sup>. However, in all cases, these movements are accompanied by a rearrangement of a cluster of conserved hydrophobic and aromatic residues ('transmission switch', made of 3.40, 5.51, 6.44 and 6.48) deeper in the receptor core<sup>58</sup>. As a result, there is a rearrangement at the TM3-TM5 interface, and formation of new non-covalent contacts at the TM5-TM6 interface. Many of the residues in this transmission switch are highly conserved in class A GPCRs, suggesting that they are likely to constitute a common feature of GPCR activation.

These agonist-induced local structural changes near the binding site are translated into larger-scale helix movements through distinct activation pathways<sup>58,59,73,74</sup>. Specifically, the changes at the TM5 bulge involving Pro<sup>5.50</sup> are transmitted through the helix, resulting in rearrangements at its cytoplasmic side<sup>75</sup>. Also, rearrangement of the TM5-TM6 interface is linked to the rotation of TM6 near Phe<sup>6.44</sup>, which is amplified owing to the strong kink of this helix, resulting in the large-scale relocation of the cytoplasmic side of TM6 (ref. 13). This opens the cleft required for binding of the G protein. This rearrangement was first demonstrated in rhodopsin using electron paramagnetic resonance spectroscopy<sup>76</sup>, and further established by fluorescence spectroscopy, ultraviolet spectroscopy and chemical cross-linking of histidines, and by double electron-electron resonance spectroscopy experiments<sup>77,78</sup>. Infrared spectroscopy on rhodopsin labelled with genetically encoded probes suggests that smaller conformational changes in TM6 in early intermediate states precede its larger rigid-body movement<sup>79</sup>. The extent of the movement of TM6 varies among receptors<sup>58</sup>, reaching 14 Å in the structure of the  $\beta_2$ -AR-G protein complex<sup>9</sup> (Fig. 5a). As discussed above, not all agonist-bound GPCR structures present these large-scale rearrangements in the cytoplasmic region, although many of them feature the rearrangement of the transmission switch<sup>58</sup>. These structures must therefore correspond either to an initial encounter complex between ligand and receptor or to 'intermediate' active states that have not manifested the full set of conformational changes to allow binding of the G protein. Thus, it seems that some receptors only adopt a fully active conformation when the G protein or other interaction partners (such as a peptide or antibody) stabilize the opening of the cytoplasmic domain. Water molecules have also been observed to facilitate conformation-specific contacts; for instance, in the inactive- and active-state rhodopsin structures, two distinct water-mediated hydrogen-bonding networks are present that involve many of the highly conserved residues of class A receptors<sup>13,14</sup>.

### TM-intracellular region changes during receptor activation

Residues from the ICL2 and cytoplasmic end of TM3 (Arg<sup>3.50</sup> of the conserved DRY motif) interact with the G protein after activation<sup>9</sup>. In rhodopsin, Arg<sup>3.50</sup> forms a salt bridge with Glu<sup>6.30</sup> in the inactive state that is broken after activation. However, this 'ionic lock' is not a conserved feature of the inactive conformation of all GPCRs and has been observed to be conformationally plastic, thus implicating this region in basal activity of receptors<sup>73</sup>. The conformation of a native ICL3, which connects TM5 and TM6, is known only in rhodopsin<sup>80</sup> and  $A_{2A}$ AR (refs 7, 12). Comparison of the structures of  $A_{2A}$ AR T4-lysozyme chimaera<sup>24</sup> with that of the thermostabilized<sup>12</sup> or antibody-bound<sup>7</sup>  $A_{2A}$ ARs shows that fusion of the lysozyme distorts the cytoplasmic sides of TM5 and TM6. Interestingly, some of the structures of  $\beta_1$ -AR with a shortened form of ICL3 are seen either with or without the ionic lock and suggests that this interaction may have a role in the regulation of receptor activation and also in basal activity<sup>81</sup>.

Comparison of the active-state structures of the  $\beta_2$ -AR bound to the G protein ( $G_i$ )<sup>9</sup> and of rhodopsin (metarhodopsin II) bound to a peptide that resembles the C-terminal tail of  $G_i$  (ref. 33) allows the identification of a consensus interface in the TM helices for G-protein binding. This is formed by at least eight positions from TM3, TM5 and TM6 (Fig. 5). To create this interface, TM5 and TM6 move considerably compared to the inactive states (Fig. 5). It should be stressed, however, that the N terminus of the G protein interacts with the ICL2 and that the structure of

ICL2 is crucial to this interaction. Hydrogen-deuterium exchange mass spectrometry showed that the receptor perturbs the structure of the amino-terminal region of the  $\alpha$ -subunit of G<sub>s</sub> and consequently alters the 'P-loop' that binds the  $\beta$ -phosphate in GDP, thereby influencing the affinity to bind GTP<sup>82</sup>. It also seems that there is allosteric communication between the G-protein interface and the ligand-binding region as it has been shown that G proteins induce or stabilize a conformation in the receptor that binds agonists with a 100-fold higher affinity<sup>83</sup>. This suggests that the allosteric coupling between the ligand-binding site and the G-protein-binding site is bidirectional<sup>7</sup>.

### Biased signalling

Although GPCRs primarily operate by coupling to G proteins, interaction with other scaffold proteins, such as  $\beta$ -arrestins, can influence signalling events<sup>84</sup>. Certain ligands can preferentially trigger some of these signalling pathways, a phenomenon designated as biased agonism or functional selectivity<sup>85</sup>. Biased ligands that can selectively influence beneficial signalling pathways may have a high therapeutic potential as drug candidates. Nevertheless, it cannot be ignored that some biased ligands might affect undesirable signalling pathways leading to drug side effects. Despite its relevance to medical and biological fields, structural and mechanistic aspects underlying biased signalling are not yet well understood. However, recent studies are beginning to provide new insights in this direction. For instance, the first structures of  $\beta_1$ -AR bound to arrestin-biased ligands<sup>36</sup> show that they have weaker interactions with TM5 than full agonists, but contact additional residues in a 'minor' binding pocket near TM7 and involving ECL2, possibly involved in arrestin-biased signalling<sup>86</sup>. These thermostabilized structures do not show large conformational changes in the cytoplasmic side of the receptor. However, the additional ligand-receptor interactions with TM7 could affect the conformational states of this helix via a conserved activation pathway through TM2 or TM7 (ref. 74). In line with this, recent <sup>19</sup>F-NMR spectroscopy experiments show that binding of biased ligands alters the local environment of a chemically modified Cys at the TM7-H8 interface<sup>87</sup>. In addition, lanthanide-based resonance energy transfer spectroscopy studies on the arginine-vasopressin type 2 receptor showed that the G-protein-biased agonists stabilize a conformation of TM6 and TM7-H8 that is distinct from that stabilized by the arrestin-biased agonists<sup>88</sup>. Taken together, these studies suggest the existence of parallel and partially decoupled pathways of receptor activation that result in G-protein and  $\beta$ -arrestin signalling. However, we are far from a molecular understanding of how these conformational changes in the receptor are translated into cellular signalling events.

### Structural features of class B, class C and other GPCRs

Although relatively little is known about the overall structure of non-class A GPCRs, important progress has been made to understand the structural and functional aspects of the extracellular ligand-binding domains of class B, class C and other receptors (covered extensively in ref. 39). In class B GPCRs, structures of N-terminal domains from different receptors have been solved by solution NMR and X-ray crystallography, revealing a common fold formed by an N-terminal  $\alpha$ -helix and two  $\beta$ -sheets stabilized by three conserved disulphide bridges<sup>89</sup>. Ligands interact with these extracellular domains to induce receptor activation, but it is not clear whether the TM bundle contributes to a shared orthosteric-binding site<sup>70</sup>. Class C GPCRs possess a large N-terminal domain containing a bilobed structure that forms the ligand-binding site. The structure of this domain (the 'venus flytrap module') has been solved by X-ray crystallography for several members of this class. Receptor activation has been proposed to proceed through allosteric coupling between the ligand-bound extracellular domain and the transmembrane bundle<sup>90</sup>. Structures of the GPCR-autoproteolysis inducing (GAIN) domain from adhesion GPCRs are beginning to provide insights into the molecular basis of receptor function for the adhesion family of GPCRs<sup>91</sup>. The availability of structures of the TM domain region for these classes of GPCRs may provide the basis for understanding whether common structural

features govern all GPCR classes. Given the limited variability in the G-protein repertoire, which is in sharp contrast to the enormous diversity in the shape and size of ligands across different GPCR classes, it is likely that the G-protein interaction interface of the receptor will be more similar than the ligand-binding site across the GPCR classes.

### Outlook and future directions

The recent development of several technologies to facilitate GPCR structure determination will probably allow elucidation of structures of virtually any GPCR encoded in the human genome, although undoubtedly some will be more tractable than the others. Thus, structures of class B receptors (for example, peptide-hormone receptors) and class C receptors (for example, the metabotropic glutamate receptor) will probably be determined soon. These GPCRs share less sequence similarity with class A receptors, making them difficult to model on the basis of existing structures. Therefore, the structures of GPCRs from these additional classes may illuminate new modes of ligand binding and signal transmission, whereas comparison with class A structures will reveal universal principles underlying GPCR signalling.

Different ligands can affect whether a GPCR couples to one or several G proteins, or preferentially to arrestin and other intracellular effectors. This raises many key questions that are yet to be addressed: what is the molecular basis of the coupling of GPCRs with different intracellular effectors? How do different ligands selectively stabilize different active conformations or influence the conformational equilibrium? What is the role of water molecules, ions, cholesterol and native lipids in GPCR structure and function? An important step towards answering these questions is to obtain more high-resolution structures of GPCRs in the fully activated state in complex with different ligands and diverse intracellular effector proteins. With the determination of structures of increased resolution and application of complementary approaches, one can gain insights into how endogenous small molecules and the near native environment of the receptor can modulate function<sup>92</sup>. Another key structural problem is to understand how GPCRs associate to form homo- or hetero-oligomers. Some crystal structures suggest potential models for how GPCR dimers could form in the cell<sup>25,27</sup>, but these need to be tested rigorously *in vivo*<sup>93</sup>. Furthermore, recent promising developments of free electron lasers will allow structure determination from small and easier to obtain nanocrystals. These advances will facilitate a better understanding of the structural determinants and functional implications of the activation mechanisms and oligomerization interfaces of receptors.

With approximately 30% of the known drugs targeting GPCRs, the pharmacological relevance of this family is firmly established. Structures of more GPCRs will facilitate structure-based *in silico* drug discovery and the development of therapeutic compounds with improved specificity and pharmacodynamics<sup>48</sup>. For example, the first preclinical compounds developed from structure-based drug design have recently been described for A<sub>2A</sub>R as candidates for the treatment of Parkinson's disease<sup>94</sup>. Although GPCRs are particularly amenable to high-throughput *in silico* screening owing to deep binding cavities where the endogenous ligands bind, recent structural studies suggest other possible regions on the receptor for rational drug development. For example, one possibility would be to modulate receptor activity at the intracellular G-protein-binding site using short cell-penetrating peptides (pepdicins) or single-domain antibodies. The structures of receptors highlight those regions that are unique to particular receptor subtypes, which may facilitate more rational developments of subtype-specific drugs, either using traditional small molecule chemistry or innovative approaches such as bicyclic or linear peptides<sup>95</sup>. Furthermore, these regions may be amenable to the design of allosteric modulators or bitopic ligands<sup>96</sup> such as small molecules, peptides or antibody domains<sup>97</sup>. Also, structure determination of GPCR mutants involved in human diseases will allow understanding of their molecular causes, and help design new therapeutic approaches for their treatment.



Although GPCR structures will be exceedingly informative, a systems-level understanding of the interactions between GPCRs and intracellular effectors in the cell will provide a holistic understanding of the cellular response under physiological and disease conditions. This will require a detailed characterization of which GPCRs and intracellular effectors are present in different cell types and, importantly, their abundance and the kinetic parameters of their interaction<sup>98,99</sup>. In addition to a systems biology approach and the discovery of new signalling functions, a genomic approach will be essential to understand the link between receptor polymorphisms in the population and how these may be linked to disease and drug efficacy. In this context, large international consortia that exploit next-generation sequencing technology such as the '1000 genome' and 'cancer genome' projects are identifying natural variants and discovering disease-causing mutations. Such receptor polymorphisms may affect activation kinetics and drug selectivity in subtle but important ways, perhaps explaining the heterogeneous response to drugs between individuals<sup>33,35,36,100</sup>. The growing number of solved structures provides a framework to now interpret such functional heterogeneity, paving the path towards personalized medicine and hence improving human health.

Received 25 June 2012; accepted 7 January 2013.

- Fredriksson, R., Lagerström, M. C., Lundin, L. G. & Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
- Tate, C. G., Schertler, G. F. & Engineering, G. Protein-coupled receptors to facilitate their structure determination. *Curr. Opin. Struct. Biol.* **19**, 386–395 (2009).
- Cherezov, V. *et al.* High-resolution crystal structure of an engineered human  $\beta_2$ -adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265 (2007).
- Rosenbaum, D. M. *et al.* GPCR engineering yields high-resolution structural insights into  $\beta_2$ -adrenergic receptor function. *Science* **318**, 1266–1273 (2007). **Description of the T4 lysozyme fusion strategy that led to the first high-resolution  $\beta_2$ -AR structure using the lipidic cubic phase crystallization technique<sup>3</sup>, which has been subsequently used to facilitate the crystallization of many other GPCRs.**
- Thompson, A. A. *et al.* Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* **485**, 395–399 (2012).
- Rasmussen, S. G. *et al.* Crystal structure of the human  $\beta_2$  adrenergic G-protein-coupled receptor. *Nature* **450**, 383–387 (2007).
- Hino, T. *et al.* G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* **482**, 237–240 (2012).
- Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the  $\beta_2$  adrenoceptor. *Nature* **469**, 175–180 (2011).
- Rasmussen, S. G. *et al.* Crystal structure of the  $\beta_2$  adrenergic receptor–G<sub>s</sub> protein complex. *Nature* **477**, 549–555 (2011). **A seminal paper showing how an activated GPCR binds to a heterotrimeric G protein, which also suggests a mechanism for how the G protein becomes activated.**
- Warne, T. *et al.* Structure of a  $\beta_1$ -adrenergic G-protein-coupled receptor. *Nature* **454**, 486–491 (2008). **This paper describes the first structure determined through the implementation of systematic scanning mutagenesis to develop a thermostabilized receptor that can be co-crystallized in the presence of even weakly binding ligands<sup>32</sup>.**
- Lebon, G. *et al.* Agonist-bound adenosine A<sub>2A</sub> receptor structures reveal common features of GPCR activation. *Nature* **474**, 521–525 (2011).
- Doré, A. S. *et al.* Structure of the adenosine A<sub>2A</sub> receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure* **19**, 1283–1293 (2011).
- Standfuss, J. *et al.* The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **471**, 656–660 (2011).
- Deupi, X. *et al.* Stabilized G protein binding site in the structure of constitutively active metarhodopsin-II. *Proc. Natl Acad. Sci. USA* **109**, 119–124 (2012).
- Caffrey, M. Crystallizing membrane proteins for structure determination: use of lipidic mesophases. *Annu. Rev. Biophys.* **38**, 29–51 (2009).
- Chae, P. S. *et al.* Maltose–neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nature Methods* **7**, 1003–1008 (2010).
- Moukhametzyanov, R. *et al.* Protein crystallography with a micrometre-sized synchrotron-radiation beam. *Acta Crystallogr. D* **64**, 158–166 (2008).
- Palczewski, K. *et al.* Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **289**, 739–745 (2000).
- Murakami, M. & Koyama, T. Crystal structure of squid rhodopsin. *Nature* **453**, 363–367 (2008).
- Haga, K. *et al.* Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **482**, 547–551 (2012).
- Kruse, A. C. *et al.* Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* **482**, 552–556 (2012).
- Shimamura, T. *et al.* Structure of the human histamine H<sub>1</sub> receptor complex with doxepin. *Nature* **475**, 65–70 (2011).
- Chien, E. Y. *et al.* Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **330**, 1091–1095 (2010).
- Jaakola, V. P. *et al.* The 2.6 angstrom crystal structure of a human A<sub>2A</sub> adenosine receptor bound to an antagonist. *Science* **322**, 1211–1217 (2008).
- Wu, B. *et al.* Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330**, 1066–1071 (2010).
- Wu, H. *et al.* Structure of the human  $\kappa$ -opioid receptor in complex with JdTrc. *Nature* **485**, 327–332 (2012).
- Manglik, A. *et al.* Crystal structure of the  $\mu$ -opioid receptor bound to a morphinan antagonist. *Nature* (2012).
- Granier, S. *et al.* Structure of the  $\delta$ -opioid receptor bound to naltrindole. *Nature* **485**, 400–404 (2012). **This paper, along with the three accompanying papers<sup>5,26,27</sup>, allows a detailed comparison of all the major opioid receptors, which is an excellent starting point for structure-based development of subtype-specific inhibitors for pain relief.**
- White, J. F. *et al.* Structure of the agonist-bound neurotensin receptor. *Nature* **490**, 508–513 (2012). **This is the first description of a GPCR bound to a peptide agonist, which shows that agonist-specific interactions in NTSR1 occur closer to the extracellular surface than observed in other receptors.**
- Zhang, C. *et al.* High-resolution crystal structure of human protease-activated receptor 1. *Nature* **492**, 387–392 (2012).
- Hanson, M. A. *et al.* Crystal structure of a lipid G protein-coupled receptor. *Science* **335**, 851–855 (2012).
- Park, S. H. *et al.* Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* **491**, 779–783 (2012).
- Choe, H. W. *et al.* Crystal structure of metarhodopsin II. *Nature* **471**, 651–655 (2011).
- Rosenbaum, D. M. *et al.* Structure and function of an irreversible agonist- $\beta_2$  adrenoceptor complex. *Nature* **469**, 236–240 (2011).
- Warne, T. *et al.* The structural basis for agonist and partial agonist action on a  $\beta_1$ -adrenergic receptor. *Nature* **469**, 241–244 (2011). **Comparison of  $\beta_1$ -AR structures bound to different agonists and partial agonists suggests reasons for their different efficacies and why inverse agonists actively inhibit the activation of  $\beta$ -ARs.**
- Warne, T., Edwards, P. C., Leslie, A. G. & Tate, C. G. Crystal structures of a stabilized  $\beta_1$ -adrenoceptor bound to the biased agonists bucindolol and carvedilol. *Structure* **20**, 841–849 (2012).
- Xu, F. *et al.* Structure of an agonist-bound human A<sub>2A</sub> adenosine receptor. *Science* **332**, 322–327 (2011). **Along with ref. 11, this paper describes the structure of A<sub>2A</sub>R in an agonist-bound active-like state that shows common features of activation compared to  $\beta_2$ -AR.**
- Scheerer, P. *et al.* Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497–502 (2008). **The structure of opsin bound to the C-terminal fragment of transducin was the first insight into how an activated receptor may associate with a G protein and, with ref. 66, was the first structure of a GPCR in an activated state.**
- Lagerström, M. C. & Schiöth, H. B. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nature Rev. Drug Discov.* **7**, 339–357 (2008).
- Unal, H. & Karnik, S. S. Domain coupling in GPCRs: the engine for induced conformational changes. *Trends Pharmacol. Sci.* **33**, 79–88 (2012).
- Hurst, D. P. *et al.* A lipid pathway for ligand binding is necessary for a cannabinoid G protein-coupled receptor. *J. Biol. Chem.* **285**, 17954–17964 (2010).
- Dror, R. O. *et al.* Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl Acad. Sci. USA* **108**, 13118–13123 (2011). **Long timescale molecular dynamics simulations are used to define a potential intermediate in the GPCR conformational change from the active to inactive state, and shows the potential for studying the activation mechanism of GPCR through molecular simulations.**
- González, A., Perez-Acle, T., Pardo, L. & Deupi, X. Molecular basis of ligand dissociation in  $\beta$ -adrenergic receptors. *PLoS ONE* **6**, e23815 (2011).
- Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).
- Madabushi, S. *et al.* Evolutionary trace of G protein-coupled receptors reveals clusters of residues that determine global and class-specific functions. *J. Biol. Chem.* **279**, 8126–8132 (2004).
- Barth, P., Wallner, B. & Baker, D. Prediction of membrane protein structures with complex topologies using limited constraints. *Proc. Natl Acad. Sci. USA* **106**, 1409–1414 (2009).
- Schlinkmann, K. M. *et al.* Critical features for biosynthesis, stability, and functionality of a G protein-coupled receptor uncovered by all-versus-all mutations. *Proc. Natl Acad. Sci. USA* **109**, 9810–9815 (2012).
- Congreve, M., Langmead, C. J., Mason, J. S. & Marshall, F. H. Progress in structure based drug design for G protein-coupled receptors. *J. Med. Chem.* **54**, 4283–4311 (2011).
- Heitz, F. *et al.* Site-directed mutagenesis of the putative human muscarinic M2 receptor binding site. *Eur. J. Pharmacol.* **380**, 183–195 (1999).
- Katritch, V., Cherezov, V. & Stevens, R. C. Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol. Sci.* **33**, 17–27 (2011).

51. Valiquette, M., Parent, S., Loisel, T. P. & Bouvier, M. Mutation of tyrosine-141 inhibits insulin-promoted tyrosine phosphorylation and increased responsiveness of the human  $\beta_2$ -adrenergic receptor. *EMBO J.* **14**, 5542–5549 (1995).
  52. Jaakola, V. P., Prilusky, J., Sussman, J. L. & Goldman, A. G. Protein-coupled receptors show unusual patterns of intrinsic unfolding. *Protein Eng. Des. Sel.* **18**, 103–110 (2005).
  53. Gsponer, J. & Babu, M. M. The rules of disorder or why disorder rules. *Prog. Biophys. Mol. Biol.* **99**, 94–103 (2009).
  54. Babu, M. M., Kriwacki, R. W. & Pappu, R. V. Structural biology. Versatility from protein disorder. *Science* **337**, 1460–1461 (2012).
  55. Qin, K., Dong, C., Wu, G. & Lambert, N. A. Inactive-state preassembly of  $G_{\alpha_s}$ -coupled receptors and  $G_q$  heterotrimers. *Nature Chem. Biol.* **7**, 740–747 (2011).
  56. Nobles, K. N. *et al.* Distinct phosphorylation sites on the  $\beta_2$ -adrenergic receptor establish a barcode that encodes differential functions of  $\beta$ -arrestin. *Sci. Signal.* **4**, ra51 (2011).
  57. Samama, P., Cotecchia, S., Costa, T. & Lefkowitz, R. J. A. Mutation-induced activated state of the  $\beta_2$ -adrenergic receptor. Extending the ternary complex model. *J. Biol. Chem.* **268**, 4625–4636 (1993).
  58. Deupi, X. & Standfuss, J. Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr. Opin. Struct. Biol.* **21**, 541–551 (2011).
  59. Hofmann, K. P. *et al.* A G protein-coupled receptor at work: the rhodopsin model. *Trends Biochem. Sci.* **34**, 540–552 (2009).
  60. Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
  61. Bockenhauer, S., Furstenberg, A., Yao, X. J., Kobilka, B. K. & Moerner, W. E. Conformational dynamics of single G protein-coupled receptors in solution. *J. Phys. Chem. B* **115**, 13328–13338 (2011).
  62. Ghanouni, P. *et al.* Functionally different agonists induce distinct conformations in the G protein coupling domain of the  $\beta_2$  adrenergic receptor. *J. Biol. Chem.* **276**, 24433–24436 (2001).
  63. Dror, R. O. *et al.* Activation mechanism of the  $\beta_2$ -adrenergic receptor. *Proc. Natl Acad. Sci. USA* **108**, 18684–18689 (2011).
  64. Kahsai, A. W. *et al.* Multiple ligand-specific conformations of the  $\beta_2$ -adrenergic receptor. *Nature Chem. Biol.* **7**, 692–700 (2011).
  65. West, G. M. *et al.* Ligand-dependent perturbation of the conformational ensemble for the GPCR  $\beta_2$  adrenergic receptor revealed by HDX. *Structure* **19**, 1424–1432 (2011).
  66. Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W. & Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183–187 (2008).
  67. Süel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature Struct. Biol.* **10**, 59–69 (2003).
  68. Vaidehi, N. & Kenakin, T. The role of conformational ensembles of seven transmembrane receptors in functional selectivity. *Curr. Opin. Pharmacol.* **10**, 775–781 (2010).
  69. Provasi, D., Artacho, M. C., Negri, A., Mobarec, J. C. & Filizola, M. Ligand-induced modulation of the free-energy landscape of G protein-coupled receptors explored by adaptive biasing techniques. *PLOS Comput. Biol.* **7**, e1002193 (2011).
  70. Wheatley, M. *et al.* Lifting the lid on GPCRs: the role of extracellular loops. *Br. J. Pharmacol.* **165**, 1688–1703 (2012).
  71. Ahuja, S. *et al.* Location of the retinal chromophore in the activated state of rhodopsin\*. *J. Biol. Chem.* **284**, 10190–10201 (2009).
  72. Bokoch, M. P. *et al.* Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* **463**, 108–112 (2010).
  73. Lebon, G., Warne, T. & Tate, C. G. Agonist-bound structures of G protein-coupled receptors. *Curr. Opin. Struct. Biol.* **22**, 482–490 (2012).
  74. Deupi, X., Standfuss, J. & Schertler, G. Conserved activation pathways in G-protein-coupled receptors. *Biochem. Soc. Trans.* **40**, 383–388 (2012).
  75. Sansuk, K. *et al.* A structural insight into the reorientation of transmembrane domains 3 and 5 during family A G protein-coupled receptor activation. *Mol. Pharmacol.* **79**, 262–269 (2011).
  76. Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L. & Khorana, H. G. Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. *Science* **274**, 768–770 (1996).
  77. Sheikh, S. P., Zvyaga, T. A., Lichtarge, O., Sakmar, T. P. & Bourne, H. R. Rhodopsin activation blocked by metal-ion-binding sites linking transmembrane helices C and F. *Nature* **383**, 347–350 (1996).
  78. Altenbach, C., Kusnetzow, A. K., Ernst, O. P., Hofmann, K. P. & Hubbell, W. L. High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. *Proc. Natl Acad. Sci. USA* **105**, 7439–7444 (2008).
  79. Ye, S. *et al.* Tracking G-protein-coupled receptor activation using genetically encoded infrared probes. *Nature* **464**, 1386–1389 (2010).
  80. Li, J., Edwards, P. C., Burghammer, M., Villa, C. & Schertler, G. F. Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol.* **343**, 1409–1438 (2004).
  81. Moukhametzianov, R. *et al.* Two distinct conformations of helix 6 observed in antagonist-bound structures of a  $\beta_1$ -adrenergic receptor. *Proc. Natl Acad. Sci. USA* **108**, 8228–8232 (2011).
  82. Chung, K. Y. *et al.* Conformational changes in the G protein  $G_s$  induced by the  $\beta_2$  adrenergic receptor. *Nature* **477**, 611–615 (2011).
  83. Whorton, M. R. *et al.* A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl Acad. Sci. USA* **104**, 7682–7687 (2007).
  84. Rajagopal, S., Rajagopal, K. & Lefkowitz, R. J. Teaching old receptors new tricks: biasing seven-transmembrane receptors. *Nature Rev. Drug Discov.* **9**, 373–386 (2010).
  85. Kenakin, T. Collateral efficacy in drug discovery: taking advantage of the good (allosteric) nature of 7TM receptors. *Trends Pharmacol. Sci.* **28**, 407–415 (2007).
  86. Rosenkilde, M. M., Benned-Jensen, T., Frimurer, T. M. & Schwartz, T. W. The minor binding pocket: a major player in 7TM receptor activation. *Trends Pharmacol. Sci.* **31**, 567–574 (2010).
  87. Liu, J. J., Horst, R., Katritch, V., Stevens, R. C. & Wuthrich, K. Biased signaling pathways in  $\beta_2$ -adrenergic receptor characterized by  $^{19}\text{F}$ -NMR. *Science* **335**, 1106–1110 (2012).
  88. Rahmeh, R. *et al.* Structural insights into biased G protein-coupled receptor signaling revealed by fluorescence spectroscopy. *Proc. Natl Acad. Sci. USA* **109**, 6733–6738 (2012).
  89. Parthier, C., Reedtz-Runge, S., Rudolph, R. & Stubbs, M. T. Passing the baton in class B GPCRs: peptide hormone activation via helix induction? *Trends Biochem. Sci.* **34**, 303–310 (2009).
  90. Urwyler, S. Allosteric modulation of family C G-protein-coupled receptors: from molecular insights to therapeutic perspectives. *Pharmacol. Rev.* **63**, 59–126 (2011).
  91. Araç, D. *et al.* A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoprolysis. *EMBO J.* **31**, 1364–1378 (2012).
  92. Liu, W. *et al.* Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **337**, 232–236 (2012).
  93. Han, Y., Moreira, I. S., Urizar, E., Weinstein, H. & Javitch, J. A. Allosteric communication between protomers of dopamine class A GPCR dimers modulates activation. *Nature Chem. Biol.* **5**, 688–695 (2009).
  94. Congreve, M. *et al.* Discovery of 1,2,4-triazine derivatives as adenosine  $A_{2A}$  antagonists using structure based drug design. *J. Med. Chem.* **55**, 1898–1903 (2012).
- This paper describes the development of new preclinical compounds for the treatment of Parkinson's disease by structure-based drug design, and shows structures of the lead compounds in the thermostabilized  $A_{2A}R$  previously developed, and the structure determined bound to inverse agonists<sup>12</sup>.**
95. Heinis, C., Rutherford, T., Freund, S. & Winter, G. Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nature Chem. Biol.* **5**, 502–507 (2009).
  96. Valant, C., Robert Lane, J., Sexton, P. M. & Christopoulos, A. The best of both worlds? Bitopic orthosteric/allosteric ligands of G protein-coupled receptors. *Annu. Rev. Pharmacol. Toxicol.* **52**, 153–178 (2012).
  97. Herr, D. R. Potential use of G protein-coupled receptor-blocking monoclonal antibodies as therapeutic agents for cancers. *Int. Rev. Cell Mol. Biol.* **297**, 45–81 (2012).
  98. Huber, T. & Sakmar, T. P. Escaping the flatlands: new approaches for studying the dynamic assembly and activation of GPCR signaling complexes. *Trends Pharmacol. Sci.* **32**, 410–419 (2011).
  99. Kiel, C. *et al.* Structural and functional protein network analyses predict novel signaling functions for rhodopsin. *Mol. Syst. Biol.* **7**, 551 (2011).
  100. Ahles, A., Rochais, F., Frambach, T., Bunemann, M. & Engelhardt, S. A. Polymorphism-specific “memory” mechanism in the  $\beta_2$ -adrenergic receptor. *Sci. Signal.* **4**, ra53 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank A. Deonarine, C. Chothia, D. Ghosal, J. Marsh, J. Garcia-Nafria, K. R. Vinothkumar, R. Henderson, R. Hegde, S. Balaji, S. Chavali and T. Flock for their comments on this work. This work was supported by the UK Medical Research Council (U105185859), HFSP (RGY0073/2010; M.M.B.), the EMBO Young Investigator Program (M.M.B.), and ERASysBio+ (GRAPPLE; M.M.B.). A.J.V. acknowledges LMB Cambridge Scholarship and St. John's College Benefactor Scholarship for financial support. G.L. was funded by Heptares Therapeutics, the UK Medical Research Council and by the CNRS and Agence Nationale de la Recherche (grant ANR-09-BLAN-0272). X.D. and G.F.S. acknowledge the Swiss National Science Foundation (grant 31003A\_132815) and the ETH Zürich within the framework of the National Center for Competence in Research in Structural Biology Program for financial support. C.G.T. acknowledges the Medical Research Council Technology Development Gap Fund, Pfizer, and core funding from the UK Medical Research Council (U105197215). We apologize to our colleagues whose work was not cited owing to space limitations.

**Author Contributions** A.J.V. and M.M.B. designed the study, analysed the results and wrote the manuscript. A.J.V. performed all calculations. X.D., G.L., C.G.T. and G.F.S. provided data, and contributed to the analysis and writing of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.J.V. ([ajv@mrc-lmb.cam.ac.uk](mailto:ajv@mrc-lmb.cam.ac.uk)) or M.M.B. ([madanm@mrc-lmb.cam.ac.uk](mailto:madanm@mrc-lmb.cam.ac.uk)).



# Thin crust as evidence for depleted mantle supporting the Marion Rise

Huaiyang Zhou<sup>1</sup> & Henry J. B. Dick<sup>2</sup>

The global ridge system is dominated by oceanic rises reflecting large variations in axial depth associated with mantle hotspots. The little-studied Marion Rise is as large as the Icelandic Rise, considering both length and depth, but has an axial rift (rather than a high) nearly its entire length. Uniquely along the Southwest Indian Ridge systematic sampling allows direct examination of crustal architecture over its full length. Here we show that, unlike the Icelandic Rise, peridotites are extensively exposed high on the rise, revealing that the crust is generally thin, and often missing, over a rifted rise. Therefore the Marion Rise must be largely an isostatic response to ancient melting events that created low-density depleted mantle beneath the Southwest Indian Ridge rather than thickened crust or a large thermal anomaly. The origin of this depleted mantle is probably the mantle emplaced into the African asthenosphere during the Karoo and Madagascar flood basalt events.

Following Morgan<sup>1</sup>, oceanic rises are thought to form via a hot fertile mantle plume producing a flow of plume-derived mantle to the ridge and down the subaxial asthenospheric channel, resulting in increased mantle temperature, ridge topography, and thickened igneous crust; this is the mantle wedge hypothesis (see ref. 2 for example). Studies of the Reykjanes ridge support this, with seismic crust thickening from about 6 km near the Gibbs fracture zone for 1,600 km to about 18 km at the Reykjanes peninsula, for example<sup>3</sup>. The Icelandic rise also has a long axial high consistent with such robust magmatism<sup>4</sup> (Table 1). The current consensus is that the igneous crust thickens over the Marion Rise, and indeed up all rises<sup>5,6</sup>. The basalt sodium contents, as at other rises, decrease systematically with proximity to the Marion hotspot, and are interpreted to represent higher degrees of mantle melting and thicker crust<sup>7</sup>. Such correlations, however, can be due to variable mantle temperature or to an increasingly depleted mantle source composition, and do not require thicker crust<sup>8</sup>. The Marion Rise, in particular, has numerous large-offset transforms that would block subaxial asthenospheric flow<sup>9</sup>, and its deep rift valley indicates weak rather than robust magmatism<sup>4</sup>. Niu and O'Hara<sup>10</sup>, although assuming thick crust over rises, suggest that the global correlation of basalt chemistry and ridge depth is best explained by mantle

composition variations, with many rises supported by depleted chemically buoyant mantle, as for the Iceland<sup>11</sup> and Azores<sup>12</sup> rises.

The only available seismic data predicts 2–6 km of crust in the Southwest Indian Ridge (SWIR) rift mountains at 57° E and 61° E (ref. 13). Modelling using sea surface gravity gives 2–4 km in the eastern SWIR, but around 6–12 km north of Marion Island<sup>14</sup>. Given that mantle peridotites are locally abundantly exposed, however, 6–12 km of igneous crust seems unlikely in the latter region. Muller *et al.*<sup>15</sup> found the Mohorovičić discontinuity (the Moho)  $5 \pm 1$  km beneath the Atlantis Bank at 57° E, but concluded that it could be a serpentinization front. Similarly, the seismic determination of crustal thickness at 61° E is at odds with mapping east of the Melville fracture zone that shows the ridge to consist of local magmatic centres linked by nearly amagmatic oblique segments producing “smooth” sea floor<sup>16</sup>. Hence, the equivalency of seismic and igneous ‘crust’ is not established for slow- and ultraslow-spreading ridges<sup>17</sup>.

All estimates of ocean crust thickness are by inference from geochemistry, seismic data and gravity. Nowhere have they been tested by direct geologic constraints, which would require deep-ocean drilling. Mapping and sampling in lower crust and mantle exposed in oceanic core complexes, however, can constrain crustal thickness<sup>18,19</sup>. Using

**Table 1 | Characteristics of rifted and axial rises**

Rise	Dominant morphology	Spreading rate (mm yr <sup>-1</sup> )	Hotspot magma flux (m <sup>3</sup> s <sup>-1</sup> )	Crustal volume (10 <sup>7</sup> km <sup>3</sup> )	Elevation range, minimum to maximum (m)	200-km-average elevation range (m)	Depth anomaly (m)	Rise length (km)	Cross-sectional area (km <sup>2</sup> )	Geographic location
Iceland	Axial high	19.5	7.0	4.53 (ref. 5)	750, –3,115	642, –2,425	3,067	2,390	3,665,156	Jan Mayen to Gibbs fracture zone at 53° S
Galapagos	Axial high	63.0	5.5 (ref. 6)	NA	–1,361, –3,732	–1,762, –3,170	1,408	1,380	970,988	Cocos–Nazca Rise 85.2°–97.5° W
Azores	Axial rift	20.5	NA	1.15 (ref. 5)	–345, –4,127	–1,659, –3,441	1,782	2,877	2,563,557	MAR 47.5°–25.8° N
Marion	Axial rift	14.6	0.4 (ref. 14)	0.95 (ref. 14)	–858, –5,150	–2,547, –4,681	2,134	3,408	3,636,236	SWIR 32.32°–63.34° E

NA, not available. The maximum and minimum elevation is measured at the lowest and the highest topographic point on each rise. These points are defined by the midpoints of the rift-valley axis where the smoothed ridge depth stops decreasing or increasing. The 200-km average maximum and minimum elevations represent 200-km along-axis average depths taken about these points to eliminate local topographic effects. The depth anomaly is computed from the average depths. Cross-sectional area is simply an estimate of size of the rise based on half its length times its total depth anomaly, which shows that the Iceland and Marion rises, in these terms, are approximately the same size.

<sup>1</sup>State Key Laboratory of Marine Geology, Tongji University, Shanghai, 200092, China. <sup>2</sup>Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

this approach, H.Z. surveyed a ridge segment at 53° E where low-basalt  $\text{Na}_8$  (ref. 20;  $\text{Na}_8$  is the Na content of basalt corrected for fractional crystallization), refractory peridotite, gravity<sup>14</sup>, and the mantle wedge hypothesis suggested that 4–8 km of crust might be present. Instead of thick crust, however, H.Z. found nearly amagmatic spreading, supporting our interpretation of sampling elsewhere on the ridge.

## Tectonic setting

The SWIR is a highly oblique ultraslow-spreading ridge with a near-constant full rate of approximately  $14 \text{ mm yr}^{-1}$  (ref. 21), extending 7,700 km between the Bouvet and Rodriguez triple junctions (Fig. 1). East of the Bouvet triple junction at 25° E it is abruptly offset about 1,800 km to the northeast by the DuToit, Andrew Bain, Marion, and Prince Edward fracture zones<sup>22</sup>. There it crosses the flank of the Southern Ocean geoid high, following a subdued positive arch over the Marion swell from the Andrew Bain fracture zone to the Gallieni fracture zone. The Marion Rise extends for around 3,400 km, starting at a depth of 4,800 m at 27° E before rising abruptly to a depth of 858 m at 36° E, and then down to 5,150 m at 63° E.

A 1.5-km-high and 70-km-long axial volcano at 36° E forms a short axial ridge marking the maximum volcanic activity on the rise. This is attributed to flow of plume-derived mantle to the ridge from beneath the Marion hotspot 256 km to the south<sup>23</sup>. Gravity, however, indicates that any influence of the Marion hotspot is limited to west of the Discovery II fracture zone<sup>9,24</sup>. Except for the 36° E segment, the entire SWIR has a deep axial rift. The SWIR east of the Indomed fracture zone formed by propagation of the Rodriguez triple junction over the last 64 million years<sup>21</sup>. It has regularly spaced non-transform discontinuities, short oblique amagmatic segments, and the Atlantis II, Novara, and Melville transforms.

## Geology

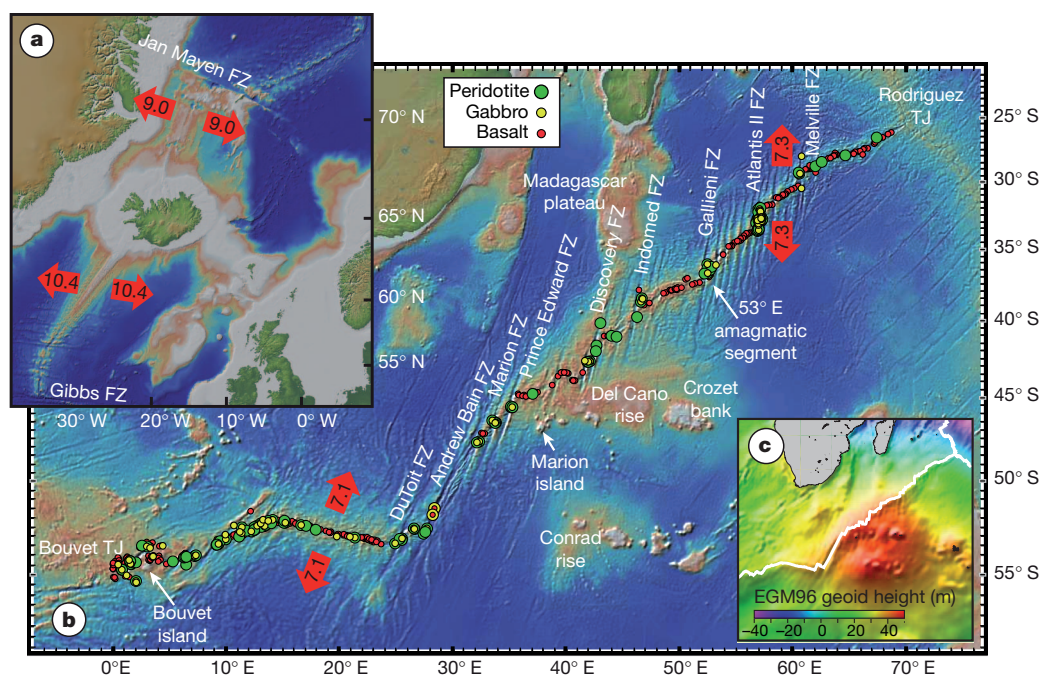
Figure 1 shows a compilation of satellite-navigated dredges (Supplementary Table 1), with which we interpreted crustal architecture along the SWIR (see Methods). Peridotite and gabbro are widely distributed over the entire ridge, although most dredging focused

on rift-valley neovolcanic zones. This severely biased recoveries to basalt, as even a thin lava veneer prevents the sampling of plutonic rock. Dredges on rift-valley walls, non-transform discontinuities, and transforms are comparatively few, except along the western SWIR and the Atlantis II fracture zone. From 9°–35° E and 41°–47° E (ref. 25) peridotite is abundant, demonstrating thin or missing basaltic crust and enormous mantle exposures<sup>25–27</sup>. Despite its frequent occurrence, however, gabbro is volumetrically scarce, and, excluding the Atlantis II fracture zone (Fig. 2b), accounts for less than 11% of the plutonic rocks dredged along the SWIR (Fig. 2a). Thus, while gabbro and peridotite occur in 25 dredges at 9°–16° E, there is 1,570 kg of peridotite but only 49 kg of gabbro. With only scattered basalt flows dredged in the rift valley, this is essentially a 400-km-long ‘amagmatic’ segment<sup>28</sup>. Similarly, although there are 21 peridotite-bearing and 17 gabbro-bearing Scripps Protea Expedition dredges from 31°–39° E, they recovered 3,451 kg of peridotite, and only 799 kg of gabbro.

Most SWIR peridotites are partially serpentinized (20%–70%) statically altered granular rocks with fine serpentine vein networks (Supplementary Figs 1 and 2). They have protogranular texture with smooth, curved grain boundaries interlocking with olivine, or an overprinting porphyroclastic texture due to high-temperature crystal-plastic deformation. Associated mylonites, which constitute around 5% of abyssal peridotites<sup>29</sup>, reflect extreme deformation produced by strain localization, with olivine–spinel geothermometry indicating closure temperatures of about 600–750 °C (ref. 30). Hydrothermally altered and sheared talc–serpentine and talc–tremolite schist represent less than 10% of the peridotites. These assemblages reflect progressive strain localization from high-temperature crystal-plastic deformation (at over 1,000 °C), followed by mylonite formation, to low-temperature brittle faulting accompanying unroofing and final emplacement to the sea floor<sup>19,31</sup>.

## The 53° E amagmatic segment

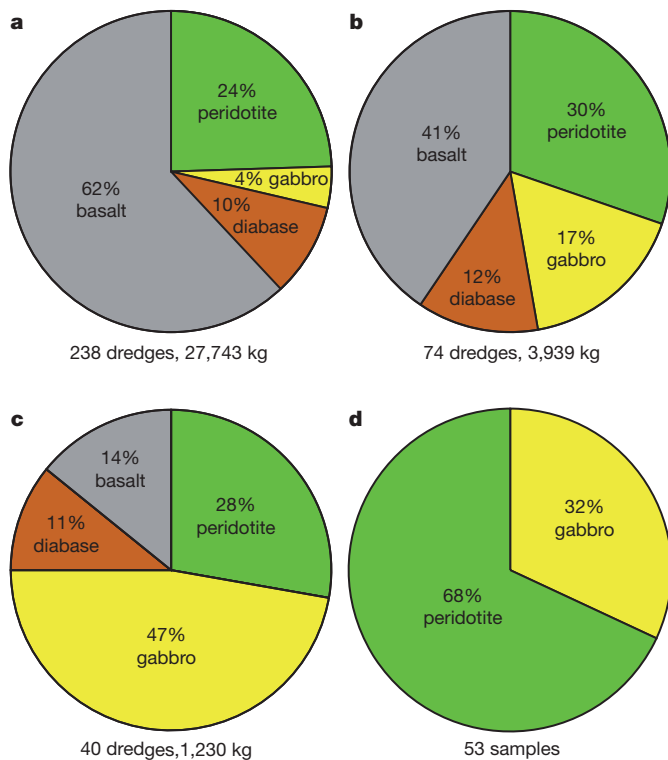
In 2010 the RV *Dayang Yihao* mapped approximately 11,000 km<sup>2</sup> between the Gallieni and Gauss fracture zones (Fig. 3). This comprised a ridge segment with a deep rift valley with 2.4-km relief at



**Figure 1 | Physiographic and geoid maps of the SWIR with sample locations, and a map of the Icelandic rise for comparison.** Spreading directions and half rates are given in millimeters on the red arrows. **a**, Relief map of the Icelandic swell. **b**, Map of the Marion swell with location of satellite-navigated dredges on the SWIR. **c**, Free-air geoid map showing the position of

the SWIR (courtesy of Mark Behn). EGM96 is Earth Gravitational Model 1996. Maps of the SWIR and Icelandic Rise plotted at the same scale for comparison. (www.geomapp.org). Spreading rates were calculated using the global plate velocity model Nuvel 1. Symbol sizes are not proportional to recovery. Data and sources are given in Supplementary Table 1. FZ, fracture zone; TJ, triple junction.



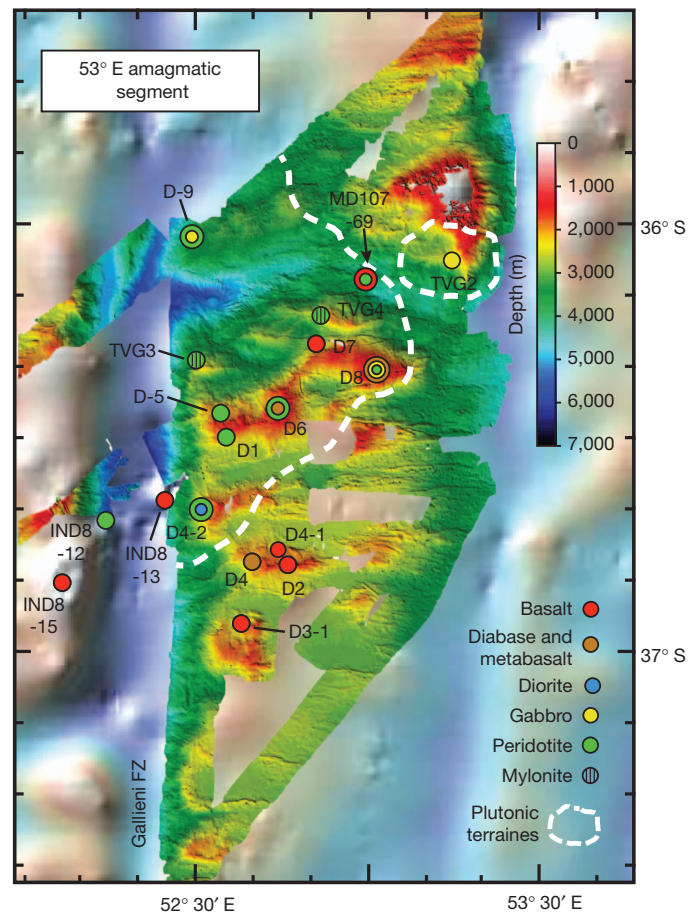


**Figure 2 | Lithologic proportions by weight for SWIR dredges compiled from Supplementary Table 1 with proportions for the MAR Atlantis massif.** **a**, SWIR excluding the Atlantis II fracture zone at 57° E. **b**, Atlantis II fracture zone including the Atlantis Bank. **c**, Atlantis Bank excluding the rest of the Atlantis II fracture zone (32.5–33° S). **d**, The Atlantis massif, MAR, with lithologic proportions by number of dive samples collected on the flanking transform wall, compiled from ref. 47. The transform gabbros are interpreted as a precursor net-vein complex intruded into massive granular peridotite from the 1,400-m gabbro sequence drilled 4.5 km to the north at Integrated Ocean Drilling Program Hole U1309D.

its midpoint, plunging from a water depth of 3.9 km to over 5 km at the 107-km-offset Gallieni fracture zone. The rift valley lacks an axial high, and is bowed north at its centre. Eleven dredges and three television-guided (TV) grabs (which spot-sample the sea floor) collected about 1,938 kg of rock (Supplementary Table 2). The Scripps Institution of Oceanography Indomed Expedition Leg 8 sampled peridotite (dredge 12) and basalt (dredge 15) from the western Gallieni Transform wall, and basalt low on the eastern wall (dredge 13), while the RV *Marion Dufresne* Edul Cruise Leg 107 dredged 2 kg of dunite and 20 kg of basalt from the rift-valley centre (dredge 69).

RV *Dayang Yihao* recovered 422 kg of basaltic rock. Diabase occurred with peridotite in dredges 6 and 8, with a small gabbro xenolith in one diabase. The remaining basalts are fine-grained to very-fine-grained, some with pillow rinds and one with glass. They are heavily weathered, but hydrothermally unaltered. Where associated with peridotite outcrops, they are probably hanging-wall debris rafted on fault footwalls. A 250-kg gabbro block from the eastern inside-corner high and seafloor physiography indicates a possible 380-km<sup>2</sup> gabbroic core complex there, consistent with volcanic crust indicated by the east-west lineated terrain in the eastern rift mountains.

Peridotite is extensively exposed in the western half of the segment and to the north and south in the rift mountains flanking the Gallieni fracture zone. The western outside-corner high lacks well developed ridge-parallel lineations, consisting instead of small domes where 204 kg of serpentinized peridotite and a small amount of gabbro were dredged. Seven stations in the southern rift mountains recovered around 1,062 kg, mostly harzburgite. Three *en echelon* domed ridges, progressively staggered to the north from the transform to the segment



**Figure 3 | Bathymetric map of the 53° E ridge amagmatic segment with sample locations and contents.** The Simrad EM120 multi-beam map covers around 11,000 km<sup>2</sup> embedded in the sea surface gravity map from GeoMapApp ([www.geomapp.org](http://www.geomapp.org)). Filled circles indicate dredge and TV grab locations. Multiple lithologies are indicated by nested symbols, with smaller insets representing the volumetrically lesser rock type. D and TVG prefixes represent the RV *Dayang Yihao* cruise 21 samples; IND indicates the Scripps Institution of Oceanography Indomed Expedition Leg 8; MD107 indicates the RV *Marion Dufresne* Cruise 107 samples. Dashed white lines outline areas of probable massive exposure of peridotite (around 3,200 km<sup>2</sup>, western area) and gabbro (around 380 km<sup>2</sup>, eastern area) interpolated using seafloor morphologies. The colour scale gives depth in metres.

centre, flank the western rift valley. They are about 8–11 km wide and 20–28 km long, shoaling to 1,000–2,800 m depth, with the largest emerging from the centre of the rift valley. TV grabs from the eastern and westernmost domed ridges recovered serpentinized granular and mylonite peridotite. Immediately south of the easternmost, dredge 8 recovered diabase, gabbro and serpentinite from a large 28 km × 13 km north–south corrugated dome. Further southwest at about 36° 28' S, peridotite was dredged at three locations on another irregular ~20 km × 35 km domed region with evidence of north–south corrugations. Peridotite with a diorite fragment was found on the transform wall at the end of an east–west ridge at approximately 36° 40' S (dredge D4-2). Outside these mantle exposures, only basalt was dredged to the south-east of D4-2, defining the limits of the mantle outcrops. The peridotite exposure flanking the Gallieni transform, then, is about 110 km in the spreading direction and up to 60 km wide, or about 3,200 km<sup>2</sup>—probably much bigger if the peridotite outcrops extend further north.

Compared to other large fracture zones, the uplift at the western inside-corner high is modest, composed of the three small axis-parallel domes, rather than one large dome, similar to those found on the opposing outside-corner high. These are distinct from more regularly lineated terrains characteristic of rifted volcanic crust in the eastern rift

mountains. Such terrain is also inconsistent with the asymmetric detachment faulting seen at large oceanic core complexes. Rather, it is generally associated with block faults exposing massive peridotite at amagmatic ridge segments<sup>16</sup>. The broader domes sampled by dredges 1, 5, 6 and 8, however, feature corrugations and some mixed lithologies characteristic of oceanic core complexes. Bathymetry and dredging, then, indicate a progressive eastward transformation from magmatic to amagmatic spreading from about 9.4 million years ago to the present in the western portion of the 53° E segment (assuming symmetric spreading), with the formation of core complexes followed by more symmetric block faulting exposing an enormous region of partially serpentinized peridotite.

Spinel in 16 peridotites from seven stations from the rift mountains and eastern Gallieni fracture-zone wall (Supplementary Table 3) and ten peridotites from two stations from the western wall (H.J.B.D., personal communication) have average spinel  $\text{Cr} \times 100/(\text{Cr} + \text{Al}) \approx 30$  corresponding to about 12% average melting<sup>32</sup> of a hypothetical primitive upper-mantle composition<sup>33</sup>. A basalt glass fragment from the 53° E segment has  $\text{Na}_8$  of 2.8; whereas those from the ridge to the east and west have average  $\text{Na}_8$  of 2.5, predicting around 4–6 km of crust and a ridge depth of around 2,500 m (ref. 20)—which is 1,000 m shallower than what is actually observed.

## Crustal architecture

Abyssal peridotites have three principal emplacement modes, each with characteristic unique textures, distribution, and associations (see Methods). These include diapiric emplacement of highly sheared talc–serpentine schists and serpentinites along both low- and high-angle faults cutting laterally or vertically through the crust into peridotite. Texturally, however, most abyssal peridotites are typical of those in large ophiolite peridotite massifs. Consistent with this, 65 of 119 SWIR peridotite dredges recovered no gabbro, and many of the rest only gabbro veins. A typical dredge covers 500–1,000 m, and multiple dredges often find little but peridotite in an area. A detachment fault exposing mantle peridotite from beneath 4–8 km of gabbro and basalt requires many kilometres of throw and heave, and gabbroic hanging-wall debris should be scattered over the fault surface along with pillow lava and diabase. There would also be gabbro inliers in the footwall owing to fault imbrication.

Where ocean drilling or mapping has found massive gabbro, as at the Atlantis Bank and the Atlantis massif, it is extensively exposed intruding peridotite on the transform wall (Fig. 2c, d). Similarly, at 23° N on the Mid-Atlantic Ridge (MAR) gabbro complexes intruded into mantle peridotite were mapped by dredging<sup>34</sup> and then confirmed and extended by seismic refraction<sup>35</sup>. A series of Ocean Drilling Program (ODP) holes in MAR tectonic windows at 14°–16° N (ref. 36) also confirm a ‘plum pudding’ model of gabbro intrusions in serpentinized mantle<sup>37</sup> constructed from bathymetry, dredging and diving. In these holes abundant gabbro veins and dikes were consistent precursors of larger gabbro bodies. The numerous gabbro veins and intrusions in peridotite on the transform wall of the Atlantis massif (Fig. 2d) then suggested the presence of the over-1,400-m gabbro body later drilled a few kilometres to the north. In all these cases, gabbro is far more abundant than it is generally along the SWIR. Moreover, gabbroic core complexes, such as the Atlantis Bank<sup>38</sup>, the Atlantis massif<sup>18</sup>, the Kane Megamullion<sup>19</sup>, and 23° S (ref. 39) all show that where gabbroic crust is present, detachment faults root there, not in the mantle. Thus, on the basis of rock textures and distribution, the SWIR peridotite localities mainly represent massive partially serpentinized mantle exposed over large seafloor regions.

Thus sampling shows thin or missing crust over the Marion Rise. Although peridotite is exposed on transform walls by detachment faulting rooted through the nearby dike–gabbro transition, as at the Atlantis II fracture zone, in most areas the associated gabbro bodies had to be very small. Many peridotite exposures, as at 53° E, are due to amagmatic spreading. Where gabbro is largely missing, the crust is

probably limited to <1–2 km, based on the known depths of the dike–gabbro transition, and hence the conductive–advective heat flow transition beneath ridges<sup>40</sup> (cooling is too rapid for gabbro to form in the advective cooling regime). In rare areas, as at Atlantis Bank, where gabbro is abundant, the crust may be over 4 km thick<sup>18</sup>, whereas in regions where largely mantle is exposed it may be largely missing. With this degree of lateral heterogeneity a precise estimate of average crustal thickness is difficult. It is possible that in the eastern SWIR, where ‘smooth’ sea floor indicative of block and detachment faulting predominates<sup>41</sup>, the average crust could be as little as 1 km, and that over the Marion swell, where regions with more igneous crust may occur, the average crust could be about 2 km. But geologically, there is not nearly enough crust along the ridge to explain the formation of the Marion Rise using any model that precludes mantle compositional buoyancy as the major contributor.

## Discussion

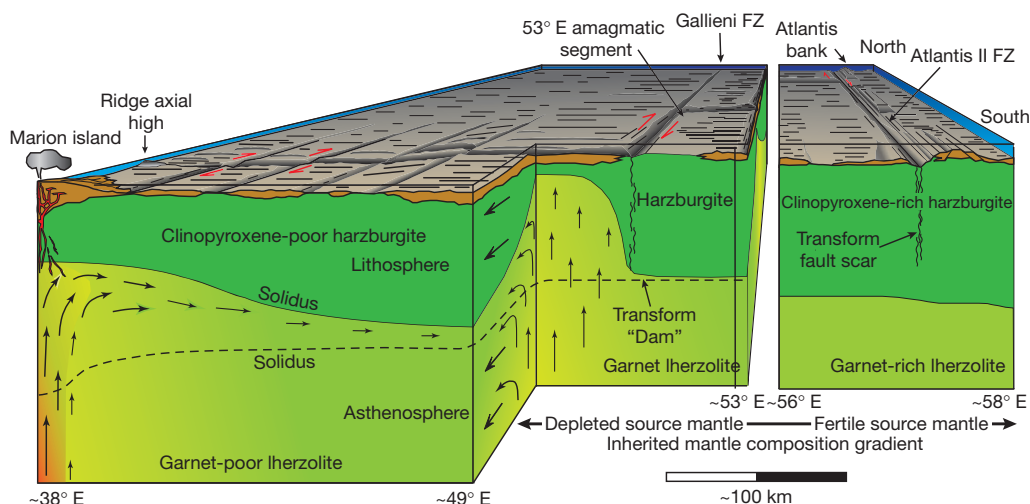
The numerous exposures of mantle peridotite and sparsity of gabbro over the Marion Rise overturns the iconic model that rises are all characterized by thick crust. This confirms the hypothesis that ocean rises can be supported by highly depleted buoyant mantle. Opposing models with uniform mantle composition and large variations in mantle potential temperature require correspondingly large variations in crustal thickness<sup>20</sup>. Differences between rifted and axial rises also cannot be attributed to spreading rate, as this is virtually uniform along the Marion Rise, and varies little up the Iceland and Azores rises, whereas the spreading rates of the rifted Marion and Azores rises bracket that of the axial Icelandic rise (Table 1).

Without a large flux of hot primitive mantle or thick crust, a rifted rise can be supported by a thermal or compositional anomaly. For the latter case, Niu and O’Hara<sup>10</sup> analysed global ridge-depth variability but found no convincing evidence for ridge temperature gradients exceeding 50°–100° K (ref. 10). They point out, however, that below about 70 km in the garnet peridotite facies, the density difference between fertile garnet lherzolite and depleted harzburgite is very large<sup>10–12,42</sup>—as much as 1.2%. Consistent with the large compositional variation of abyssal peridotites<sup>8</sup>, a 0.6% density contrast can easily account for the global ridge-depth variation, requiring compensation down to 458 km. Proportionately then, the local 2.1-km Marion Rise depth anomaly could be compensated in the upper 200–250 km entirely in the asthenosphere. Even assuming large temperature anomalies, however, it has been noted<sup>43</sup> that the Hawaiian swell, possibly the largest thermal plume on the planet, has a peak amplitude of only about 1.4 km. Similarly, Ito *et al.*<sup>44</sup> found that only 1.3 km of their predicted 3.7-km Icelandic topographic anomaly could be supported by a 180 °C mantle temperature anomaly, the rest being an isostatic response to crustal thickness. Thus, it is very unlikely that the geriatric Marion plume could account for the 2.1-km depth anomaly along the Marion Rise.

Mantle composition gradients along rises can explain their gravity signal, which, when modelled assuming uniform mantle composition, predicts 8–10 km of crust on the Marion and Azores rises<sup>14,45</sup>. Varying serpentinization depth is not a likely alternative because serpentine is unlikely to be stable to a depth of 9–10 km beneath a ridge. Models for the Moho as a serpentinization front attribute this to tectonic uplift into the rift mountains as the lithosphere cools into the serpentine stability field<sup>46</sup>. Where the Moho is probably a serpentinization front at the Gakkal ridge, the seismic crust is only 1.9–3.3 km thick<sup>17</sup>. Such a front is controlled by the thermal structure at the ridge axis, and if anything, its contribution to the depth anomaly should be negative if mantle temperature and crustal thickness increase towards the high.

We have shown that a large region of previously depleted mantle supports the Marion Rise rather than down-axis intrusion of hot primitive mantle (Fig. 4). Thin crust requires this explanation, because any thermal anomaly requires a more refractory mantle to offset the accompanying increase in melting. Without an enormous





**Figure 4 | Cartoon looking north from the Antarctic plate across the SWIR, showing the inferred crustal architecture along the Marion Rise.** Note that transforms near Marion Island are not shown with the proper offsets to emphasize how a large-offset transform offsets the lithosphere and blocks sub-axial asthenospheric flow<sup>48</sup>. Right side of the cartoon illustrates the Atlantis

Bank oceanic core complex, which is shown on the front face of the diagram, though its actual position (indicated by arrow) is to the north, flanking the transform valley. The ridge section between 53° E and 56° E with several oblique amagmatic rifts<sup>49</sup> is not shown.

thermal anomaly, mantle density variations seem to be the only possible explanation. The SWIR formed with the breakup of Gondwanaland about 180 million years ago, and the asthenosphere beneath the ridge was drawn up from beneath southern Africa, Madagascar and Antarctica. Therefore, it probably represents the depleted source mantle for the Karoo, Madagascar, and Ferrar flood basalt events emplaced into the asthenosphere. If this mantle represents compositionally buoyant mantle plumes, the depletion may go back even further in Earth history.

The other major rifted rise—the Azores Rise—is similar to the Marion Rise. It has only a 95-km-long axial high at its crest, and a deep axial rift along the rest of its 2,900-km length. Despite a focus on the neovolcanic zone, as at the Marion Rise, almost accidental sampling has found peridotites exposed at numerous locations up its length (Supplementary Fig. 3). In conclusion, then, there are two classes of rise: axial and rifted, one robustly magmatic, the other weakly so. Although their associated hotspots and tracks indicate related origins, the dramatic differences between them require explanation. We propose that, in general, the robustly magmatic axial rises are largely supported by present-day melting anomalies, and the weakly magmatic rifted rises are largely supported by mantle residues from the earlier history of the associated plumes.

## METHODS SUMMARY

To interpret the crustal architecture along the SWIR, we collected as much detailed information on the proportions, state of alteration, tectonic and igneous fabrics, and primary igneous petrology of our rock samples as we could. This data was then used to identify the probable manner in which these samples were exposed to the sea floor, whether emplaced as large massifs by detachment faulting, or serpentine diapirism up high-angle faults that cut either vertically or laterally into mantle rock<sup>50</sup>. We took particular care to exclude talc-serpentine and talc-tremolite schists and highly sheared serpentinites when assessing the proportion of mantle rocks collected from the sea floor that represent major basement outcrops.

Critical to our methodology are mantle and lower-crustal rocks exposed by detachment faults at oceanic spreading centres. These faults often operate for millions of years, exposing peridotite and gabbro on their footwalls, while the volcanic carapace spreads in the opposite direction. This produces oceanic core complexes that can expose many hundreds of square kilometres of plutonic rock on the sea floor, providing sections into the deep crust and upper mantle of regional significance. Their discovery and exploration over the past 20 years has revolutionized our understanding of crustal accretion at slow- and ultra-slow-spreading ocean ridges. A further development, which we discuss, is the discovery of large areas along the SWIR where there is little igneous crust, and mantle rock spreads onto the sea floor in both directions by extensional block

faulting at amagmatic accretionary ridge segments. Thus, interpretation of the data used here is a product of many years of exploration of plutonic rocks outcropping on the sea floor, as well as ocean drilling deep sections into lower crust and mantle massifs, which has led to an understanding of the mechanisms by which they are emplaced and their tectonic significance.

**Full Methods** and any associated references are available in the online version of the paper.

Received 30 September 2011; accepted 11 December 2012.

Published online 6 February 2013.

- Morgan, W. J. Deep mantle convection plumes and plate motions. *Am. Assoc. Petrol. Geol. Bull.* **56**, 203–213 (1972).
- Kincaid, C., Schilling, J.-G. & Gable, C. The dynamics of off-axis plume-ridge interaction in the uppermost mantle. *Earth Planet. Sci. Lett.* **137**, 29–43 (1996).
- Weir, R. W. *et al.* Crustal structure of the northern Reykjanes Ridge and Reykjanes Peninsula, southwest Iceland. *J. Geophys. Res.* **106**, 6347–6368 (2001).
- Ito, G. & Behn, M. D. Magmatic and tectonic extension at mid-ocean ridges: 2. Origin of axial morphology. *Geochem. Geophys. Geosyst.* **9**, Q09012, doi:10.1029/2008GC001970 (2008).
- Wang, T., Lin, J., Tucholke, B. & Chen, Y. J. Crustal thickness anomalies in the North Atlantic Ocean basin from gravity analysis. *Geochem. Geophys. Geosyst.* **12**, Q0AE02, doi:10.1029/2010GC003402 (2011).
- Canales, J. P., Ito, G., Detrick, R. S. & Sinton, J. Crustal thickness along the western Galapagos Spreading Center and the compensation of the Galapagos hotspot swell. *Earth Planet. Sci. Lett.* **203**, 311–327, doi:10.1016/S0012-821X(02)00843-9 (2002).
- Cannat, M., Rommevaux-Jestin, C., Sauter, D., Deplus, C. & Mendel, V. Formation of the axial relief at the very slow spreading Southwest Indian Ridge (49° to 69° E). *J. Geophys. Res.* **104**, 22825–22843 (1999).
- Dick, H. J. B., Fisher, R. L. & Bryan, W. B. Mineralogic variability of the uppermost mantle along mid-ocean ridges. *Earth Planet. Sci. Lett.* **69**, 88–106 (1984).
- Georgen, J. E., Lin, J. & Dick, H. J. B. Models of mantle upwelling beneath the Southwest Indian Ridge: the effects of ridge-transform geometry on magma supply at an ultra-slow spreading ridge. *Eos* **79**, abstr. 854 (1998).
- Niu, Y. & O'Hara, M. J. Global correlations of ocean ridge basalt chemistry with axial depth: a new perspective. *J. Petrol.* **49**, 633–664 (2008).
- O'Hara, M. J. Is there an Icelandic mantle plume? *Nature* **253**, 708–710 (1975).
- Presnall, D. C. & Helsley, C. E. Diapirism of depleted peridotite—a model for the origin of hot spots. *Phys. Earth Planet. Inter.* **29**, 148–160 (1982).
- Muller, M. R., Minshull, T. A. & White, R. S. Segmentation and melt supply at the Southwest Indian Ridge. *Geology* **27**, 867–870 (1999).
- Zhang, T., Lin, J. & Gao, J. Y. Interactions between hotspots and the Southwest Indian Ridge during the last 90 Ma: implications on the formation of oceanic plateaus and intra-plate seamounts. *Sci. China Earth Sci.* **54**, doi:10.1007/s11430-011-4219-9 (2011).
- Muller, M. R., Robinson, C. J., Minshull, T. A., White, R. S. & Bickle, M. J. Thin crust beneath ocean drilling program borehole 735B at the Southwest Indian Ridge? *Earth Planet. Sci. Lett.* **148**, 93–107 (1997).
- Cannat, M. *et al.* Modes of seafloor generation at a melt-poor ultraslow-spreading ridge. *Geology* **34**, 605–608 (2006).

17. Jokat, W. *et al.* Geophysical evidence for reduced melt production on the super-slow Gakkel Ridge (Arctic Ocean). *Nature* **423**, 962–965 (2003).
18. Dick, H. J. B. *et al.* A long in-situ section of the lower ocean crust: results of ODP Leg 176 drilling at the Southwest Indian Ridge. *Earth Planet. Sci. Lett.* **179**, 31–51 (2000).
19. Dick, H. J. B., Tivey, M. A. & Tucholke, B. E. Plutonic foundation of a slow-spreading ridge segment: oceanic core complex at Kane Megamullion, 23°30'N, 45°20'W. *Geochem. Geophys. Geosyst.* **9**, 44, doi:10.1029/2007GC001645 (2008).
20. Klein, E. M. & Langmuir, C. H. Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *J. Geophys. Res.* **92**, 8089–8115 (1987).
21. Patriat, P. & Segoufin, J. Reconstruction of the Central Indian Ocean. *Tectonophysics* **155**, 211–234 (1988).
22. Schlater, J. G., Grindlay, N. R., Madsen, J. A. & Rommevaux-Jestin, C. Tectonic interpretation of the Andrew Bain transform fault: Southwest Indian Ocean. *Geochem. Geophys. Geosyst.* **6**, doi:10.1029/2005GC000951 (2005).
23. Duncan, R. A. Hot spots in the southern oceans—an absolute frame of reference for the motion of the Gondwana continents. *Tectonophysics* **74**, 29–42 (1981).
24. Georgen, J. E., Lin, J. & Dick, H. J. B. Evidence from gravity anomalies for interactions of the Marion and Bouvet hotspots with the Southwest Indian Ridge: effect of transform offsets. *Earth Planet. Sci. Lett.* **187**, 283–300 (2001).
25. Dick, H. J. B., Lin, J. & Schouten, H. An ultraslow-spreading class of ocean ridge. *Nature* **426**, 405–412 (2003).
26. Fisher, R. L. & Dick, H. J. B. Natland, J. & Meyer, P. S. Mafic/ultramafic suites of the slowly spreading Southwest Indian Ridge: PROTEA Exploration of the Antarctic Plate Boundary, 24°E–47°E. *Ophiolite* **11**, 147–178 (1986).
27. Le Roex, A. P., Dick, H. J. B. & Fisher, R. L. Petrology and geochemistry of MORB from 25°E to 46°E along the Southwest Indian Ridge: evidence for contrasting styles of mantle enrichment. *J. Petrol.* **30**, 947–986 (1989).
28. Dick, H. J. *et al.* How variable slow-spread ocean crust. *Eos* **84**, abstr. #V22F-01 (2003).
29. Dick, H. J. B. in *Magmatism in the Ocean Basins* (eds Saunders, A. D. & Norry, M. J.) 71–105 (Geological Society Special Publication No. 42, 1989).
30. Jaroslow, G. E., Hirth, G. & Dick, H. J. B. Abyssal peridotite mylonites: implications for grain-size sensitive flow and strain localization in the oceanic lithosphere. *Tectonophysics* **256**, 17–37 (1996).
31. Ildefonse, B. *et al.* Oceanic core complexes and crustal accretion at slow-spreading ridges. *Geology* **35**, 623–626 (2007).
32. Hellebrand, E., Snow, J. E., Dick, H. J. B. & Hofmann, H. Coupled major and trace-element indicators in mid-ocean ridge peridotites. *Nature* **410**, 677–681 (2001).
33. Sun, S.-s. Chemical composition and origin of the Earth's primitive mantle. *Geochim. Cosmochim. Acta* **46**, 179–192 (1982).
34. Dick, H. J., Tivey, M. A., Tucholke, B. E. & Cheadle, M. J. The plutonic foundation of a MAR ridge spreading segment: the Kane Oceanic Core Complex. *Eos* **86**, abstr. T33G-02 (2005).
35. Xu, M., Canales, J. P., Tucholke, B. E. & DuBois, D. L. Heterogeneous seismic velocity structure of the upper lithosphere at Kane oceanic core complex, Mid-Atlantic Ridge. *Geochem. Geophys. Geosyst.* **10**, doi:10.1029/2009GC002586 (2009).
36. Kelemen, P. B. *et al.* Drilling mantle peridotite along the Mid-Atlantic Ridge from 14° to 16°N. *Proc. ODP Init. Rep.* **209**, (Ocean Drilling Program, 2004).
37. Cannat, M. *et al.* Ultramafic and gabbroic exposures at the Mid-Atlantic Ridge: geologic mapping in the 15°N region. *Tectonophysics* **279**, 193–213 (1997).
38. Blackman, D. K. *et al.* *Proceedings of the Integrated ODP 304/305* (Integrated Ocean Drilling Program Management International, doi:10.2204/iodp.proc.304305.2006 (2006)).
39. Morishita, T. *et al.* Igneous, alteration and exhumation processes recorded in abyssal peridotites and related fault rocks from an oceanic core complex along the Central Indian Ridge. *J. Petrol.* **50**, 1299–1325 (2009).
40. Teagle, D. A. *et al.* Drilling a complete in situ section of upper oceanic crust formed at a superfast spreading rate: hole 1256D. *Eos* **87** (52), abstr. B31B–1090 (2006).
41. Cannat, M., Sauter, D., Escartin, J., Lavier, L. & Picazo, S. Oceanic corrugated surfaces and the strength of the axial lithosphere at slow spreading ridges. *Earth Planet. Sci. Lett.* **288**, 174–183 (2009).
42. Jordan, T. H. in *The Mantle Sample: Inclusions in Kimberlites and Other Volcanics* (Proceedings of the Second International Kimberlite Conference) (eds Boyd, F. R. & Meyer, H. O. A.) Vol. 2, 1–14 (American Geophysical Union, 1979).
43. Cadio, C., Ballmer, M. D., Panet, I., Diamant, M. & Ribe, N. New constraints on the origin of the Hawaiian swell from wavelet analysis of the geoid to topography ratio. *Earth Planet. Sci. Lett.* **359–360**, 40–54, doi:10.1016/j.epsl.2012.10.006 (2012).
44. Ito, G., Shen, Y., Hirth, G. & Wolfe, C. J. Mantle flow, melting, and dehydration of the Iceland mantle plume. *Earth Planet. Sci. Lett.* **165**, 81–96 (1999).
45. Detrick, R. S., Needham, H. D. & Renard, V. Gravity anomalies and crustal thickness variations along the Mid-Atlantic Ridge between 33°N and 40°N. *J. Geophys. Res.* **100**, 3767–3787 (1995).
46. Canales, J. P., Detrick, R. S., Lin, J., Collins, J. A. & Toomey, D. R. Crustal and upper mantle seismic structure beneath the rift mountains and across a non-transform offset at the Mid-Atlantic Ridge. *J. Geophys. Res.* **105**, 2699–2720 (2000).
47. Karson, J. A. *et al.* Detachment shear zone of the Atlantis Massif core complex, Mid-Atlantic Ridge, 30°N. *Geochem. Geophys. Geosyst.* **7**, Q06016, doi:10.1029/2005GC001109 (2006).
48. Georgen, J. E. & Lin, J. Plume-transform interactions at ultra-slow spreading rates: Implications for the Southwest Indian Ridge. *Geochem. Geophys. Geosyst.* **4** (9), 9106, doi:10.1029/2003GC000542 (2003).
49. Dick, H. J., Lin, J., Michael, P. J., Schouten, H. & Snow, J. E. Ultra-slow-spreading—a new class of ocean ridge. *Eos* **83**, abstr. T52E-05 (2002).
50. Dick, H. J. B., Arai, S., Hirth, G., John, B. J., KROO-06 Scientific Party. A subhorizontal cross-section through the crust mantle boundary at the SW Indian Ridge. *Geophys. Res. Abstr.* **3**, 794 (2001).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by the Chinese National Key Basic Research Program (2012CB417300), China Ocean Mineral Resources Research and Development Association, and the US National Science Foundation (grant OCE-0526905). We thank the crew and scientists of RV *Dayang Yihao* Cruise 21. M. Sulanowska provided technical support. Z. Chen and Y. Liu analysed chrome spinels for us at the Guangzhou Institute of Geochemistry, Chinese Academy of Sciences. A glass sample from 53°E was analysed by F. Ji at the State Key Laboratory for Mineral Deposits Research, Nanjing University. We thank the Core and Rock Storage Facility of the Woods Hole Oceanographic Institution for curatorial support and access to samples. D. Sauter, R. Fisher and E. Bonatti provided additional unpublished sample descriptions and locations for the eastern and western SWIR. A review by Y. Niu encouraged us to provide more appropriate measurements of ridge depth and directed us to several important papers, greatly benefiting the manuscript and its conclusions.

**Author Contributions** This article represents the first report of results of a survey during Cruise 21 Leg 5 of the RV *Dayang Yihao* directed by H.Z., who also had additional analytical work done at his institution. H.J.B.D. collected the bulk of the sample data over a period of 35 years. Both authors interpreted the results, with H.J.B.D. primarily responsible for the written text.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.J.B.D. (hdick@who.edu) or H.Z. (zhouhy@tongji.edu.cn).



## METHODS

The interpretation of crustal and upper mantle structure and architecture using seafloor samples is based on an understanding of the physical features of the samples: their igneous, metamorphic and structural characteristics, the tectonic setting in which they were collected, and the local seafloor physiography. Major features, such as detachment faults, high-angle normal faults, and transform walls provide exposures of the lower crust and mantle that can easily be misinterpreted. Here we summarize many of the key characteristics that guided our interpretation of the samples collected along the SWIR.

Although they comprise a relatively small proportion of mantle-derived rocks sampled along the SWIR, talc–tremolite and talc–serpentine schists can result in misinterpretation of basement lithology. Such schists represent a small percentage of SW Indian Ridge peridotites, certainly less than 10% (Supplementary Fig. 1), but accounting for these is important to our methodology for interpreting crustal architecture. Before the Integrated Ocean Drilling Program drilling at the Atlantis massif, the presence of such rocks suggested that the basement there was massive peridotite. However, 1,400 m, largely of gabbro, was drilled instead. Evidently the schists and included mylonite fragments found on the massif<sup>47</sup> were intruded along the detachment fault from where it cut peridotite emplaced in the transform domain to where it rooted in gabbro. These schists also occur *in situ* at Atlantis Bank<sup>50</sup>, lying on 850 m of massive gabbro outcrop, and at Kane Megamullion on the MAR as fault gouge on statically altered massive peridotite and dynamically deformed gabbro<sup>19</sup>. They represent skins several metres thick, originally intruded along, and then exposed on, the footwalls of detachment faults, and are excluded in our interpretation of crustal architecture along the SWIR.

Another important consideration is serpentine diapirism up high-angle normal faults that cut into mantle rock at depth. This is seen in the Klamath Mountains of the western USA, where peridotite hydrated to serpentinite (studied by H.J.B.D.) is emplaced diapirically along high-angle faults that cut vertically or laterally into peridotite. This produces serpentinite belts up to hundreds of metres wide extending up to many tens of kilometres between unrelated metavolcanics and greywackes. They are highly deformed and sheared, often with facoidal peridotite blocks preserving primary granular textures. Missing are talc–serpentine schists, entrained mylonites, and any evidence of high-temperature faulting. This mode, representing late faulting and hydration of peridotite above the brittle–ductile transition, can account for the serpentinites in-faulted into lavas and gabbros in several shallow Deep Sea Drilling Program<sup>51</sup> and Ocean Drilling Program<sup>52</sup> MAR holes on the Azores Rise. Consequently, while the occurrence of highly sheared serpentinites support the idea that the crust may be thin, they do not provide direct evidence that the crust is missing.

Critical to our methodology is mantle and lower crust interpreted as emplaced to the sea floor by detachment faults where these rocks spread in one direction as the volcanic carapace spreads in the other<sup>53</sup>. These faults often operate for millions of

years<sup>54,55</sup>, requiring enormous heave and throw, producing distinctive bathymetric terrain<sup>16,25,56</sup>. Almost half the MAR sea floor from 12.5° N to 35° N formed by such asymmetric spreading<sup>58</sup>. At ultraslow-spreading ridges, peridotite is also emplaced over large areas by more symmetric block faulting at amagmatic ridge segments where the lithosphere is too strong to sustain low-angle faults<sup>16,25,58,59</sup>. The associated high-temperature mylonites show both these fault classes rooted at high temperatures in the brittle–ductile transition in the mantle. Dredging, diving and drilling have documented many such large mantle massifs, such as at the MAR from 14° N to 16° N (refs 36, 37, 60, 61), at the Kane Megamullion<sup>19</sup>, and on the SWIR<sup>25,62</sup>, as we also report here for the SWIR at 53° E.

Thus, interpretation of the rock sample data used in this paper to decipher crustal architecture along the SWIR is a result of many years of exploration of plutonic rocks outcropping on the sea floor by dredging, submersible and remotely operated vehicle (ROV), as well as ocean drilling deep sections into lower-crust and mantle massifs, which has led to an understanding of the mechanisms by which they are emplaced.

51. Bougault, H. & Cande, S. C. 1. Background, objectives, and summary of principal results: Deep Sea Drilling Sites 556–564. *Init. Rep. DSDP* **82**, 5–16 (1985).
52. Aumento, F. & Melson, W. G. *Initial Reports of the Deep Sea Drilling Project* Vol. 37, 1008 (US Government Printing Office, 1977).
53. Dick, H. J. B., Bryan, W. B. & Thompson, G. Low-angle faulting and steady-state emplacement of plutonic rocks at ridge-transform intersections. *Eos* **62**, 406 (1981).
54. Tucholke, B. E. & Lin, J. A geological model for the structure of ridge segments in slow spreading ocean crust. *J. Geophys. Res.* **99**, 11937–11958 (1994).
55. Cann, J. R. *et al.* Corrugated slip surfaces formed at ridge–transform intersections on the Mid-Atlantic Ridge. *Nature* **385**, 329–332 (1997).
56. Smith, D. K., Cann, J. R. & Escartin, J. Widespread active detachment faulting and core complex formation near 13° N on the Mid-Atlantic Ridge. *Nature* **442**, 440–443 (2006).
57. Escartin, J. *et al.* Central role of detachment faults in accretion of slow-spreading oceanic lithosphere. *Nature* **455**, 790–794 (2008).
58. Michael, P. J. *et al.* Magmatic and amagmatic seafloor spreading at the slowest mid-ocean ridge: Gakkel Ridge, Arctic Ocean. *Nature* **423**, 956–961 (2003).
59. Sauter, D. *et al.* Focused magmatism versus amagmatic spreading along the ultraslow spreading Southwest Indian Ridge: evidence from TOBI side scan sonar imagery. *Geochem. Geophys. Geosyst.* **5**, doi:10.1029/2004GC000738 (2004).
60. MacLeod, C. J. *et al.* Life cycle of oceanic core complexes. *Earth Planet. Sci. Lett.* **287**, 333–344 (2009).
61. Schroeder, T. *et al.* Nonvolcanic seafloor spreading and corner-flow rotation accommodated by extensional faulting at 15° N on the Mid-Atlantic Ridge: a structural synthesis of ODP Leg 209. *Geochem. Geophys. Geosyst.* **8**, doi:10.1029/2006GC001567 (2007).
62. Standish, J. J., Dick, H. J. B., Michael, P. J., Melson, W. G. & O'Hearn, T. MORB generation beneath the ultraslow-spreading Southwest Indian Ridge (9°–25° E): major element chemistry and the importance of process versus source. *Geochem. Geophys. Geosyst.* **9**, Q05004, doi:10.1029/2008GC001959 (2008).

# Identification of a candidate therapeutic autophagy-inducing peptide

Sanae Shoji-Kawata<sup>1,2</sup>, Rhea Sumpter Jr<sup>1,2</sup>, Matthew Leveno<sup>1,2</sup>, Grant R. Campbell<sup>3,4</sup>, Zhongju Zou<sup>1,2,5</sup>, Lisa Kinch<sup>5,6</sup>, Angela D. Wilkins<sup>7</sup>, Qihua Sun<sup>1,2</sup>, Kathrin Pallauf<sup>1</sup>, Donna MacDuff<sup>8</sup>, Carlos Huerta<sup>6†</sup>, Herbert W. Virgin<sup>8</sup>, J. Bernd Helms<sup>9</sup>, Ruud Eerland<sup>9</sup>, Sharon A. Tooze<sup>10</sup>, Ramnik Xavier<sup>11,12,13</sup>, Deborah J. Lenschow<sup>8,14</sup>, Ai Yamamoto<sup>15</sup>, David King<sup>16</sup>, Olivier Lichtarge<sup>7</sup>, Nick V. Grishin<sup>5,6</sup>, Stephen A. Spector<sup>3,4</sup>, Dora V. Kaloyanova<sup>9</sup> & Beth Levine<sup>1,2,5,17</sup>

**The lysosomal degradation pathway of autophagy has a crucial role in defence against infection, neurodegenerative disorders, cancer and ageing. Accordingly, agents that induce autophagy may have broad therapeutic applications. One approach to developing such agents is to exploit autophagy manipulation strategies used by microbial virulence factors. Here we show that a peptide, Tat-beclin 1—derived from a region of the autophagy protein, beclin 1, which binds human immunodeficiency virus (HIV)–1 Nef—is a potent inducer of autophagy, and interacts with a newly identified negative regulator of autophagy, GAPR-1 (also called GLIPR2). Tat-beclin 1 decreases the accumulation of polyglutamine expansion protein aggregates and the replication of several pathogens (including HIV-1) *in vitro*, and reduces mortality in mice infected with chikungunya or West Nile virus. Thus, through the characterization of a domain of beclin 1 that interacts with HIV-1 Nef, we have developed an autophagy-inducing peptide that has potential efficacy in the treatment of human diseases.**

Autophagy functions in metazoans in cellular and tissue homeostasis, physiology, development, and protection against disease, and abnormalities in autophagy may contribute to many different pathophysiological conditions<sup>1,2</sup>. Thus, strategies that augment autophagy may prevent or treat human disease<sup>3</sup>. Although some drugs in clinical use are capable of augmenting autophagy, these compounds exert pleiotropic effects, revealing an unmet need to develop specific inducers of autophagy.

We sought to develop a specific autophagy-inducing agent with a potentially wide range of therapeutic effects. As viruses often provide key insights into the functionally important domains of host proteins, we investigated the molecular determinants governing the interaction between beclin 1, an essential autophagy protein in the class III phosphatidylinositol-3-OH kinase (PI(3)K) complex involved in autophagic vesicle nucleation<sup>4</sup>, and the HIV-1 virulence factor, Nef<sup>5</sup>. These investigations led us to identify a Nef-interacting sequence of beclin 1 that is necessary and sufficient for autophagy induction and which provided the basis for the development of an autophagy-inducing peptide drug that has benefits in the clearance of polyglutamine expansion protein aggregates and the treatment of infectious diseases.

## The HIV-1 Nef-interacting domain of beclin 1

HIV-1 Nef acts as an antiautophagic maturation factor through interaction with beclin 1 (ref. 5) and is required for efficient viral replication and AIDS pathogenicity. To map the Nef-interacting domain of beclin 1, we co-transfected Flag-tagged beclin 1 deletion mutants (amino acids 1–377, 141–450 and 257–450) with haemagglutinin

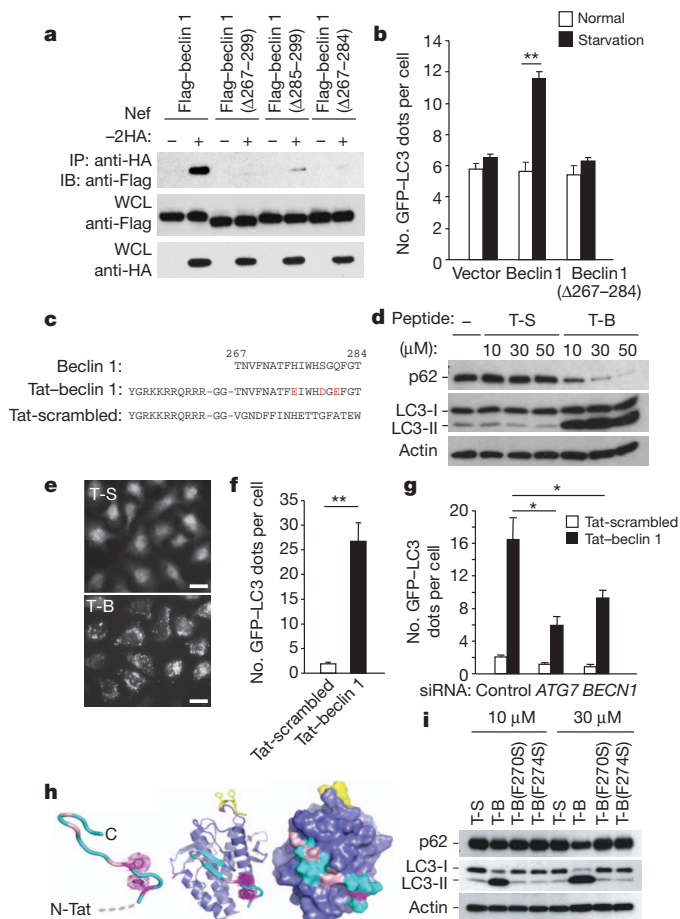
(HA)-tagged Nef. These Flag-beclin 1 mutants immunoprecipitated with Nef-HA (Supplementary Fig. 1a, b), indicating that amino acids 257–337, a region in the beclin 1 evolutionarily conserved domain (ECD)<sup>6,7</sup>, are involved in Nef binding. Studies with beclin 1 mutants containing deletions in the ECD revealed that amino acids 267–299 were required for binding to Nef (Supplementary Fig. 1c). Deletion of amino acids 267–284 weakened binding to Nef more than deletion of amino acids 285–299 (Fig. 1a) but had no effect on binding to Vps34 (also called PIK3C3) (Supplementary Fig. 2a), the class III PI(3)K that interacts with the beclin 1 ECD<sup>6</sup>. Thus, beclin 1 amino acids 267–284 in the ECD are crucial for beclin 1 binding to Nef. These residues are also required for the autophagy function of beclin 1; Flag-beclin 1, but not Flag-beclin 1( $\Delta$ 267–284), increased starvation-induced autophagy in MCF7 cells (that express low amounts of endogenous beclin 1 (ref. 8)) as measured by quantification of GFP-LC3 dots, a marker of autophagosomes<sup>9</sup> (Fig. 1b).

## Tat-beclin 1 is an autophagy-inducing peptide

We hypothesized that amino acids 267–284 of beclin 1 may be sufficient to induce autophagy, so we designed a cell-permeable peptide, Tat-beclin 1, composed of the HIV-1 Tat protein transduction domain (PTD)<sup>10</sup> attached via a diglycine linker to 18 amino acids derived from amino acids 267–284 of beclin 1 (Fig. 1c). Three substitutions were made (in residues not conserved in all Atg6/beclin 1 orthologues, Supplementary Fig. 3) to enhance peptide solubility. These substitutions had no effect on binding of full-length beclin 1 to Nef (Supplementary Fig. 1d). The 18 amino acids derived from beclin 1 in

<sup>1</sup>Department of Internal Medicine, UT Southwestern Medical Center, Dallas, Texas 75390, USA. <sup>2</sup>Center for Autophagy Research, UT Southwestern Medical Center, Dallas, Texas 75390, USA. <sup>3</sup>Department of Pediatrics, University of California, San Diego, La Jolla, California 92093, USA. <sup>4</sup>Rady Children's Hospital, San Diego, California 92123, USA. <sup>5</sup>Howard Hughes Medical Institute, UT Southwestern Medical Center, Dallas, Texas 75390, USA. <sup>6</sup>Department of Biochemistry, UT Southwestern Medical Center, Dallas, Texas 75390, USA. <sup>7</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>8</sup>Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>9</sup>Department of Biochemistry and Cell Biology and Institute of Biomembranes, Utrecht University, Utrecht 3508TD, The Netherlands. <sup>10</sup>London Research Institute, Cancer Research UK, London EC1V 4AD, UK. <sup>11</sup>Center for Computational and Integrative Biology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. <sup>12</sup>Gastrointestinal Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. <sup>13</sup>Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. <sup>14</sup>Department of Medicine, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>15</sup>Department of Neurology, Columbia University College of Physicians & Surgeons, New York, New York 10032, USA. <sup>16</sup>Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA. <sup>17</sup>Department of Microbiology, UT Southwestern Medical Center, Dallas, Texas 75390, USA. †Present address: Reata Pharmaceuticals, Inc, Irving, Texas 75063, USA.





**Figure 1 | Tat-beclin 1 peptide induces autophagy *in vitro*.**

**a**, Immunoprecipitation of Flag-beclin 1 constructs with Nef-HA in HeLa cells 24 h post-transfection. **b**, GFP-LC3-positive dots (autophagosomes) in MCF7 cells expressing GFP-LC3 and Flag-beclin 1 constructs grown in either normal medium or starved in EBSS for 2 h. **c**, Sequences of beclin 1 amino acids 267–284, Tat-beclin 1 (T-B) and Tat-scrambled (T-S) control peptide. Red letters indicate amino acid substitutions to enhance hydrophilicity. **d**, Biochemical assessment of autophagy (p62 and LC3 immunoblots) in peptide-treated HeLa cells (3 h). **e**, **f**, Representative images (**e**) and quantification of GFP-LC3-positive dots (**f**) in peptide-treated HeLa/GFP-LC3 cells (30  $\mu$ M, 3 h). Scale bars, 20  $\mu$ m. **g**, GFP-LC3-positive dots in siRNA-transfected peptide-treated HeLa/GFP-LC3 cells (30  $\mu$ M, 3 h). **h**, Model of Tat-beclin 1 peptide (left) based on corresponding elements of the beclin 1 evolutionarily conserved domain (ECD) structure (centre). Essential phenylalanine side chains, magenta; positions of solubility mutations, pink; lipid interaction site, yellow. ECD surface representation (right) illustrates exposure of corresponding peptide (cyan). **i**, p62 and LC3 immunoblots in peptide-treated HeLa cells (3 h). In **b**, **f**, **g**, bars represent mean  $\pm$  s.e.m. of triplicate samples (50–100 cells per sample). Similar results were observed in three independent experiments. \* $P < 0.05$ , \*\* $P < 0.01$ ; *t*-test.

Tat-beclin 1 were randomly shuffled to generate a control peptide (Tat-scrambled, Fig. 1c). Circular dichroism spectra of Tat-beclin 1 and Tat-scrambled revealed that both peptides are random coils (Supplementary Fig. 4), which is consistent with corresponding structural elements of Tat<sup>11</sup> and the beclin 1 ECD<sup>7</sup>.

We used several complementary methods<sup>9</sup> to demonstrate that the Tat-beclin 1 peptide induces autophagy. Treatment with Tat-beclin 1, but not Tat-scrambled, resulted in a dose-dependent decrease in amounts of p62, a selective autophagy substrate, and a dose-dependent conversion of the non-lipidated form of LC3, LC3-I, to the lipidated, autophagosome-associated form of LC3, LC3-II, in multiple cell lines and primary murine embryonic fibroblasts (MEFs) (Fig. 1d and Supplementary Fig. 5). A peptide consisting of the HIV

Tat PTD fused to the wild-type sequence of beclin 1 amino acids 267–284 was not soluble in saline-based solutions; however, when dissolved in H<sub>2</sub>O, it induced LC3-II conversion and p62 degradation more efficiently than Tat-beclin 1 (Supplementary Fig. 6). The previously reported Tat-vFLIP  $\alpha$ 2 peptide<sup>12</sup>, derived from viral-encoded FLICE inhibitory protein, induced autophagy as efficiently as Tat-beclin 1 (Supplementary Fig. 7). In HeLa/GFP-LC3 cells<sup>13</sup>, Tat-beclin 1 increased GFP-LC3 dot numbers (autophagosomes) 27-fold (Fig. 1e, f). Ultrastructurally, increased numbers of autophagosomes and autolysosomes were observed in Tat-beclin 1-treated HeLa cells (Supplementary Fig. 8a). Bafilomycin A1, an inhibitor of autophagosome/lysosomal fusion, further increased amounts of LC3-II and p62 in Tat-beclin 1-treated HeLa and COS-7 cells, indicating enhanced autophagic flux (Supplementary Fig. 8b). The magnitude of change in p62 upon bafilomycin A1 treatment was similar in Tat-beclin 1-treated cells to that observed in cells subjected to starvation (Supplementary Fig. 9). Long-lived protein degradation was increased in Tat-beclin 1-treated cells, and partially blocked by 3-methyladenine (3-MA), an autophagy inhibitor that targets class III PI(3)K (ref. 3) (Supplementary Fig. 8c).

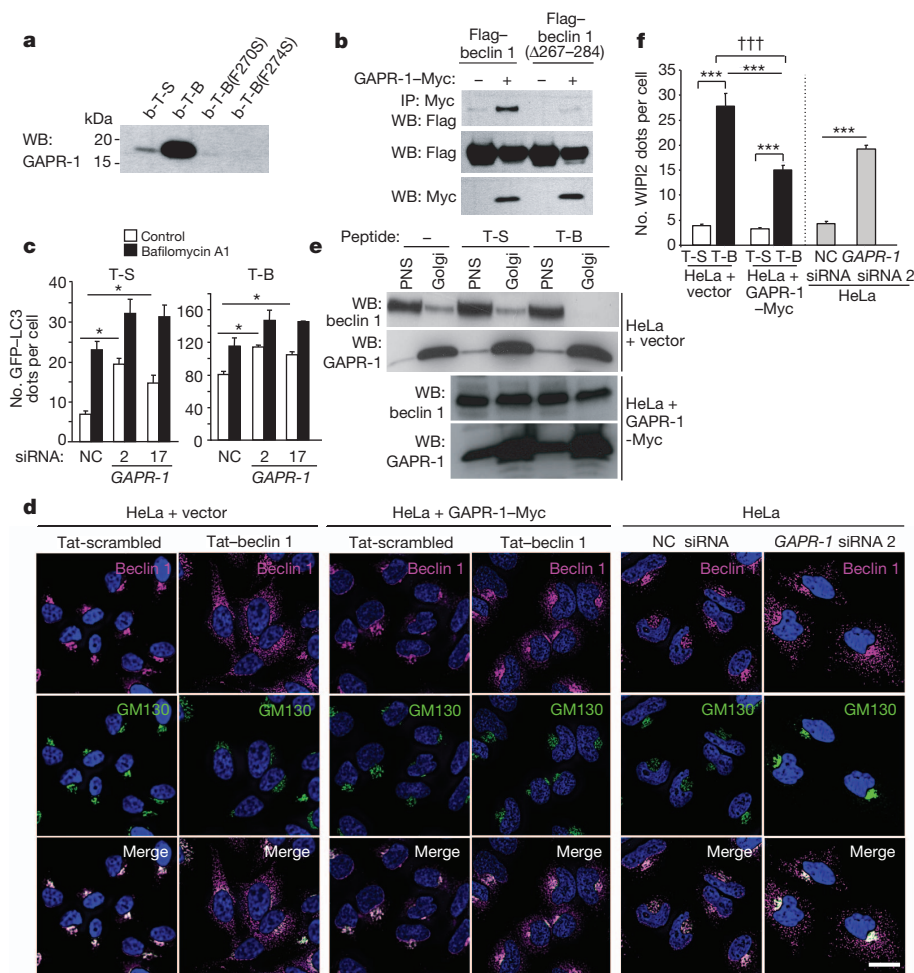
Together, these results indicate that the Tat-beclin 1 peptide induces a complete cellular autophagy response. This response involves the canonical autophagy pathway, as short interfering RNA (siRNA) knockdown of two essential autophagy genes, *BECN1* (involved in vesicle nucleation) and *ATG7* (involved in the protein conjugation system) (Supplementary Fig. 10), decreased the number of Tat-beclin 1-induced autophagosomes (Fig. 1g). The upstream signalling events in Tat-beclin 1-induced autophagy might be partially distinct from those in starvation-induced autophagy, as additive effects were observed in cells simultaneously treated with Tat-beclin 1 and subjected to starvation (Supplementary Fig. 11).

A sequence alignment of Atg6/beclin 1 orthologues revealed two conserved aromatic residues within amino acids 267–284: F270 and F274 (Supplementary Fig. 3). In the beclin 1 ECD structure (a domain that binds to liposomes *in vitro*<sup>7</sup> at the yellow-coloured site in Fig. 1h), these aromatic side chains intercalate in the hydrophobic core in a T-shaped geometry and help to position the intervening residues in a surface-exposed loop. They are predicted to have a similar function in the Tat-beclin 1 peptide, thus contributing to its structural stability (Fig. 1h). Consistent with this prediction, mutation of either of these two residues to serine (F270S and F274S) abrogated the peptide's autophagy-inducing activity (Fig. 1i). Moreover, the subcellular localization of Tat-beclin 1(F270S) lacked the discrete punctate appearance of Tat-beclin 1 (Supplementary Fig. 12).

### Tat-beclin 1 binds the autophagy inhibitor GAPR-1

We performed biochemical analyses to identify cellular proteins that bind to a biotin-conjugated form of the autophagy-inducing Tat-beclin 1 peptide. A protein that migrated at a molecular mass between 15 and 20 kDa bound to biotin-Tat-beclin 1 but not biotin-Tat-scrambled (Supplementary Fig. 13). This band was identified by liquid chromatography-tandem mass spectrometry (LC-MS/MS) as Golgi-associated plant pathogenesis-related protein 1 (GAPR-1) (also known as GLIPR2), a protein that associates with lipid rafts at the cytosolic leaflet of the Golgi membrane<sup>14</sup>. Endogenous GAPR-1, but not Vps34, immunoprecipitated with the biotin-Tat-beclin 1 peptide (Fig. 2a and Supplementary Fig. 2b). Flag-beclin 1, but not Flag-beclin 1( $\Delta$ 267–284), immunoprecipitated with GAPR-1 in HeLa cells expressing GAPR-1-Myc and Flag-beclin 1 (Fig. 2b). Thus, amino acids 267–284 of beclin 1 are necessary and sufficient for beclin 1 to bind GAPR-1. Biotin conjugates of mutant peptides defective in autophagy induction (Fig. 1i) failed to bind to GAPR-1 (Fig. 2a).

To evaluate the function of GAPR-1 in autophagy, we treated HeLa/GFP-LC3 cells with siRNAs directed against *GAPR-1* (Supplementary Fig. 14a). Knockdown of GAPR-1 led to increased numbers of autophagosomes in both Tat-scrambled- and Tat-beclin 1-treated



**Figure 2 | Tat-beclin 1 peptide binds to GPR-1, a beclin 1-interacting protein.** **a**, HeLa cells were treated with biotin-conjugated peptides (30 μM, 3 h) and proteins bound to peptides were analysed by immunoblot with anti-GAPR-1. b-T-B, biotin-Tat-beclin 1; b-T-S, biotin-Tat-scrambled. **b**, Immunoprecipitation of Flag-beclin 1 with GPR-1-Myc in HeLa cells 24 h post-transfection. **c**, GFP-LC3-positive dots in GPR-1 siRNA-transfected peptide-treated HeLa/GFP-LC3 cells (20 μM, 3 h) with or without 100 nM bafilomycin A1. Bars represent mean ± s.e.m. of triplicate samples (50–100 cells per sample). Similar results were observed in three independent experiments. NC, non-silencing control. **d**, Localization of beclin 1 and GM130

(a Golgi marker) in HeLa cells stably transduced with empty vector (left panel) or GPR-1-Myc (middle panel) or transfected with GPR-1 siRNA (right panel) and treated with peptide (20 μM, 1 h). Scale bar, 20 μm. **e**, Immunoblot of beclin 1 in post-nuclear supernatant (PNS) and Golgi-enriched fractions in HeLa cells stably transduced with empty vector or GPR-1-Myc after peptide treatment (20 μM, 2 h). **f**, WIPI2 dots in cells in the experimental conditions shown in **d**. Bars represent mean ± s.e.m. for 100–150 cells. \*P < 0.05, \*\*\*P < 0.001; t-test. †††P < 0.001; two-way ANOVA for comparison of magnitude of changes between groups.

cells (Fig. 2c). This was due to an increase in autophagic flux, rather than a block in autophagosome maturation, as numbers of autophagosomes increased further upon bafilomycin A1 treatment. (We infer that the number of lysosomes available for autophagolysosomal fusion may be rate-limiting, as absolute numbers of autolysosomes did not increase (Supplementary Fig. 15).) Taken together, our data indicate that GPR-1 is a beclin 1-interacting protein that negatively regulates autophagy.

We examined the subcellular localization of endogenous beclin 1 after Tat-beclin 1 treatment, GPR-1 overexpression or GPR-1 knockdown (Fig. 2d and Supplementary Fig. 14). Beclin 1 levels were not affected by GPR-1 overexpression or GPR-1 knockdown. In control-transfected HeLa cells treated with control peptide, beclin 1 staining predominantly overlapped with a Golgi apparatus marker (GM130). After Tat-beclin 1 peptide treatment, beclin 1 redistributed to non-Golgi subcellular compartments; beclin 1 demonstrated colabelling with markers for organelles such as endosomes (EEA1), lysosomes (LAMP1) and endoplasmic reticulum (PD1) (data not shown). In addition, we observed a significant increase in the number of punctate structures containing WIPI2 (Fig. 2f and Supplementary

Fig. 16), a mammalian protein that binds to phosphatidylinositol 3-phosphate at the site of early autophagosomal structures<sup>15</sup>. In HeLa cells expressing GPR-1-Myc, Tat-beclin 1 treatment resulted in less redistribution of beclin 1 to non-Golgi sites and decreased WIPI2 puncta. Conversely, even in the absence of Tat-beclin 1 peptide, GPR-1 siRNA knockdown led to an increase in beclin 1 redistribution to non-Golgi sites and in WIPI2 puncta.

To confirm that Tat-beclin 1 induced beclin 1 localization to non-Golgi sites, we measured the amounts of beclin 1 in purified HeLa cell Golgi fractions after peptide treatment (Fig. 2e). In control-transfected cells, treatment with Tat-beclin 1 led to the disappearance of beclin 1 in the Golgi fraction, whereas in cells expressing GPR-1-Myc, increased amounts of beclin 1 were detected in the Golgi fraction, which were only minimally decreased by treatment with Tat-beclin 1. Thus, GPR-1 may function to tether beclin 1 in the Golgi apparatus (where it is inactive in autophagy), and the Tat-beclin 1 peptide may promote the release of beclin 1 from the Golgi, resulting in enhanced early autophagosome formation. Other unknown mechanisms may also contribute to the autophagy-inducing activity of Tat-beclin 1.



## Tat-beclin 1 has beneficial effects *in vitro*

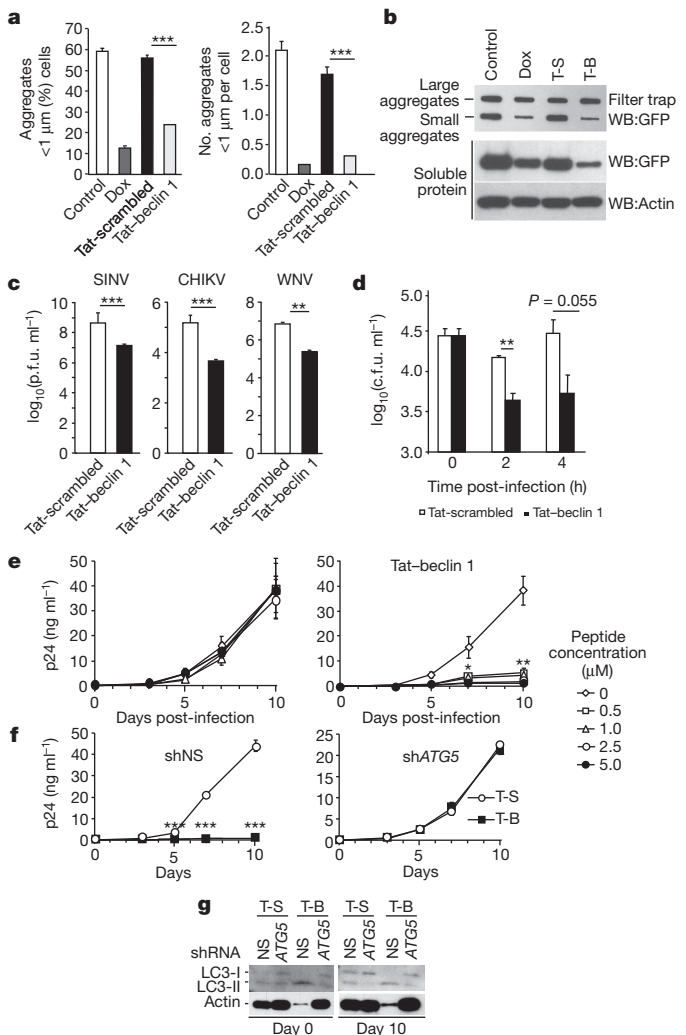
Our discovery that Tat-beclin 1 is an inducer of autophagy led us to explore whether it may represent a candidate therapeutic agent for neurodegenerative diseases and infectious diseases. Pharmacological activation of autophagy reduces levels of soluble and aggregated proteins involved in polyglutamine repeat expansion disorders (such as Huntington's disease, the spinocerebellar ataxias, and synucleinopathies and tauopathies) and reduces their cytotoxicity *in vitro* and neurotoxicity in mouse or *Drosophila* models<sup>3,16</sup>. We therefore investigated whether Tat-beclin 1 peptide treatment decreases aggregates of a model polyglutamine expansion protein (htt103Q) derived from exon 1 of human mutant huntingtin protein<sup>17</sup>.

In HeLa cells expressing doxycycline-repressible CFP fused to htt103Q (ref. 17), Tat-beclin 1 treatment did not affect the number of large (>1 µm) CFP-htt103Q aggregates (data not shown), consistent with previous reports that autophagy does not clear large protein aggregates<sup>16</sup>. In contrast, Tat-beclin 1 treatment decreased the number of small (<1 µm) CFP-htt103Q aggregates as efficiently as doxycycline (Fig. 3a and Supplementary Fig. 17). (Similar results were obtained using Torin1, an inhibitor of mTOR that induces autophagy<sup>3</sup> (Supplementary Fig. 18).) We confirmed our findings biochemically using a filter trap assay that separates large protein aggregates, small protein aggregates and soluble proteins; Tat-beclin 1 decreased small htt103Q aggregates to a similar degree as doxycycline (Fig. 3b). Tat-beclin 1 peptide also decreased amounts of soluble htt103Q protein, indicating that Tat-beclin 1 can clear the pre-aggregated and small aggregated form of htt103Q protein. Further studies are required to determine whether Tat-beclin 1 peptide is effective in animal models of full-length disease-causing polyglutamine expansion proteins.

Autophagy genes have beneficial effects in animal models of certain viral and intracellular bacterial infections, and autophagy-inducing agents (such as rapamycin and vitamin D) inhibit HIV replication in primary human monocyte-derived macrophages (MDMs)<sup>3</sup>. We therefore investigated the effects of the Tat-beclin 1 peptide on the replication of three positive-stranded RNA viruses (Sindbis virus (SINV), chikungunya virus (CHIKV), West Nile virus (WNV)), HIV-1, and the intracellular bacterium, *Listeria monocytogenes*. Tat-beclin 1 treatment (10 µM for 4 h beginning 4 h post-infection) reduced SINV, CHIKV and WNV titres 10–50-fold in HeLa cells (Fig. 3c). Tat-beclin 1 treatment reduced CHIKV replication even if administered 24 h post-infection (Supplementary Fig. 19). Tat-beclin 1-mediated reduction in viral titres was not due to peptide cytotoxicity, as the peptide did not kill uninfected cells (Supplementary Fig. 20a) and there was increased, rather than decreased, cell survival in Tat-beclin 1 peptide-treated virally infected cells (data not shown). Thus, the decreased viral titres probably reflect xenophagic degradation of the viruses or other antiviral effects of increased autophagy.

Tat-beclin 1 peptide decreased the intracellular survival of *L. monocytogenes* in primary murine bone-marrow-derived macrophages (BMDMs) (Fig. 3d). We used a strain of *L. monocytogenes* lacking the autophagy evasion protein, ActA<sup>18</sup>, and a dose of the peptide (10 µM for 2 h) that is non-toxic to uninfected BMDMs (Supplementary Fig. 20b, c). This antibacterial effect was decreased in BMDMs derived from *Atg5<sup>fllox/fllox</sup>* lysosome-M-Cre mice<sup>19</sup> that had decreased *Atg5* expression and decreased Tat-beclin 1-induced autophagy (Supplementary Fig. 21).

Tat-beclin 1 markedly inhibited HIV-1 replication in primary human MDMs. Using an established pre-treatment model<sup>20,21</sup>, we observed a dose-dependent inhibition of HIV p24 antigen release in MDMs cultured in the presence of non-toxic concentrations (0.5, 1, 2.5, and 5 µM) of Tat-beclin 1, with nearly undetectable antigen levels in cells treated daily with 5 µM Tat-beclin 1 peptide (Fig. 3e and Supplementary Fig. 22). The magnitude of inhibition of HIV-1 replication was similar to that observed with the mTOR inhibitor rapamycin, which also induces autophagy (Supplementary Fig. 23). Tat-beclin 1 peptide-mediated inhibition of HIV-1 replication



**Figure 3 | Tat-beclin 1 peptide decreases aggregates of a polyglutamine expansion protein and has anti-infective activity.** **a**, Percentage of cells with small htt103Q aggregates (left) and number of aggregates per cell (right) in HeLa cells expressing doxycycline (Dox)-repressible CFP-htt103Q after daily treatment with doxycycline or peptide (20 µM, 4 h per day) for 2 days. Bars represent mean  $\pm$  s.e.m. of triplicate samples (60–120 cells per sample). Similar results were observed in three independent experiments. **b**, Filter trap assays for htt103Q large and small aggregates in HeLa/htt103Q cells. **c**, Viral titres in HeLa cells infected with 0.1 plaque-forming units (p.f.u.) per cell of SINV, CHIKV or WNV (strain TX02) and treated with peptide (10 µM, 4–8 h post-infection). Values represent geometric mean  $\pm$  s.e.m. for triplicate samples of supernatants collected 18 h post-infection (SINV) or 24 h post-infection (CHIKV and WNV). Similar results were observed in three independent experiments. **d**, Bacteria colony-forming units (c.f.u.) in primary BMDMs infected with *L. monocytogenes* ΔactA mutant strain DPL-4029<sup>27</sup> for 30 min and treated with peptide (10 µM from 0 to 2 h post-infection). Bars represent mean  $\pm$  s.e.m. of triplicate samples. Similar results were observed in three independent experiments. **e**, HIV-1 p24 antigen release in primary human MDMs infected with HIV-1 24 h after initiation of daily peptide treatment. Values represent mean  $\pm$  s.e.m. of triplicate samples. Similar results were observed in MDMs from three independent donors. **f**, HIV-1 p24 antigen release in MDMs transduced with nonspecific scrambled shRNA (shNS) or *ATG5* shRNA (shATG5) and treated daily with peptide (5 µM). Values represent mean  $\pm$  s.e.m. of triplicate wells. Similar results were observed in MDMs from two independent donors. **g**, LC3 immunoblot of MDMs transduced with the indicated shRNA at day 0 and day 10 after HIV-1 infection. \* $P$  < 0.05; \*\* $P$  < 0.01; \*\*\* $P$  < 0.001; *t*-test.

appeared to occur via the canonical autophagy pathway, as short hairpin RNA (shRNA) knockdown of the essential autophagy gene, *ATG5* (Supplementary Fig. 22c), decreased Tat-beclin 1-induced

autophagy (as assessed by LC3 lipidation (Fig. 3g)) and completely abrogated the antiviral effects of Tat-beclin 1 treatment (Fig. 3f) without any decrease in cell survival (Supplementary Fig. 22d).

### Tat-beclin 1 induces autophagy *in vivo*

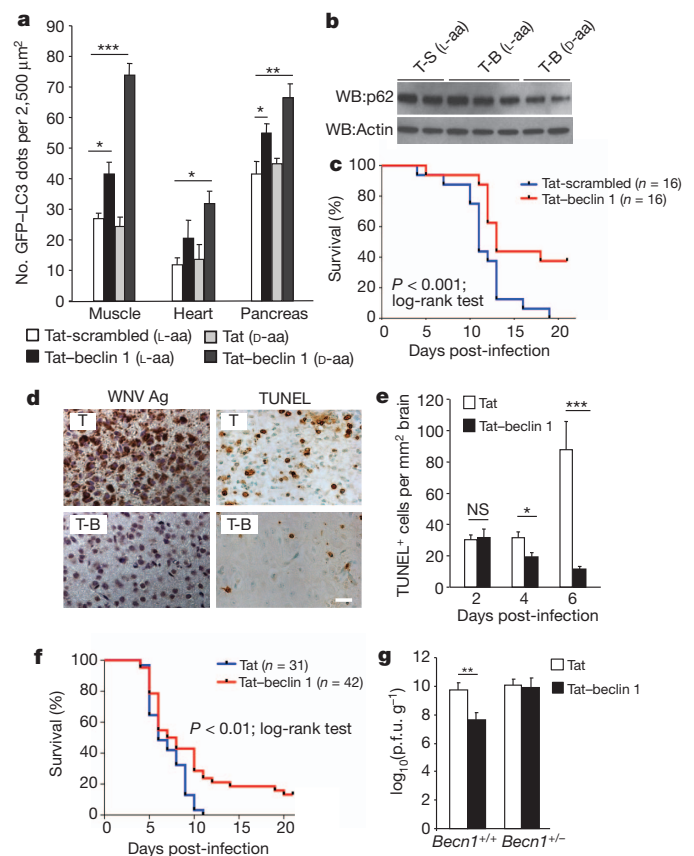
To show that Tat-beclin 1 induces autophagy *in vivo*, we examined numbers of autophagosomes in tissues of mice that transgenically express GFP-LC3 (ref. 22) and compared two forms of Tat-beclin 1, including the sequence in Fig. 1c composed of L-amino acids, and a retro-inverso sequence of Tat-beclin 1 composed of D-amino acids, which is predicted to be more resistant to proteolytic degradation<sup>23</sup>. Significantly more GFP-LC3 dots (autophagosomes) were observed in skeletal muscle, cardiac muscle and pancreas after intraperitoneal (i.p.) treatment with the L- or D-form of Tat-beclin 1 as compared to the L-form of Tat-scrambled or D-form of Tat (Fig. 4a and Supplementary Fig. 24). By imaging tissue stained with streptavidin linked to an infrared dye, we confirmed that biotin-conjugated Tat-beclin 1 enters muscle cells (Supplementary Fig. 25). In 5-day-old suckling mouse brains, treatment with the D-form (but not L-form) of Tat-beclin 1 reduced p62 levels (Fig. 4b). Thus, Tat-beclin 1 can induce autophagy in peripheral tissues in adult mice as well as in the central nervous system of neonatal mice.

Daily Tat-beclin 1 peptide treatment was well-tolerated in both neonatal and adult mice. During a 2-week course, neonatal mice had normal weight gain, motor activity, renal function and clinical status as well as normal histology of the heart, kidney, brain and liver (Supplementary Fig. 26 and Supplementary Table 1, and data not shown). These parameters, as well as hepatic and haematological function, were also normal in Tat-beclin 1 peptide-treated 3-month-old mice (Supplementary Table 1 and data not shown).

### Tat-beclin 1 improves clinical outcomes

We investigated the effect of the Tat-beclin 1 peptide on the outcome of neonatal infection with CHIKV and WNV. After subcutaneous (s.c.) inoculation with CHIKV, neonatal mice develop lethal infection similar to that observed in human neonates, characterized by flaccid paralysis and abundant viral replication in muscle, skin and joints<sup>24</sup>. Tat-beclin 1 peptide treatment (15 mg kg<sup>-1</sup> i.p. daily, L-form, beginning one day post-infection) reduced viral titres at day 4 and day 6 in the soleus and vastus lateralis muscles (Supplementary Fig. 27a) and decreased clinical paralysis (data not shown). In neonatal GFP-LC3 transgenic mice, autophagosomes were increased in muscle, including in cells that expressed CHIKV E2 envelope glycoprotein (Supplementary Fig. 27b). Importantly, Tat-beclin 1 treatment reduced mortality of neonatal mice infected with CHIKV; 100% of mice treated with control Tat-scrambled peptide died as compared to 62.5% of mice treated with Tat-beclin 1 peptide (Fig. 4c).

Arbovirus infections of the central nervous system (CNS) in human infants are often lethal and have no effective treatment<sup>25,26</sup>. We evaluated the effects of the Tat-beclin 1 peptide (D-form) on clinical outcome in a neonatal mouse model of WNV CNS infection. Tat-beclin 1 peptide treatment increased brain viral titres at day 2 post-infection, but resulted in more than a 1,000-fold reduction in brain viral titres by day 6 post-infection (Supplementary Fig. 28). The brains of control Tat peptide (D-form)-treated mice had higher levels of viral antigen staining at day 4 and day 6 post-infection, more neuropathology, and more cell death than those of Tat-beclin 1 peptide-treated mice (Fig. 4d, e and data not shown). Tat-beclin 1 peptide treatment also significantly reduced mortality in WNV-infected mice (Fig. 4f). (The antiviral effects of Tat-beclin 1 *in vivo* may be autophagy-dependent, as CNS viral titres were not reduced in Tat-beclin 1 peptide-treated mice with monoallelic deletion of *Becn1* (Fig. 4g).) Taken together, our findings demonstrate that the Tat-beclin 1 peptide improves the clinical outcome of mice with CNS WNV infection.



**Figure 4 | Tat-beclin 1 peptide induces autophagy and exerts antiviral activity *in vivo*.** **a**, GFP-LC3-positive dots in tissues of 6-week-old GFP-LC3 mice treated with the indicated peptide (20 mg kg<sup>-1</sup> i.p., 6 h). A minimum of ten fields was counted per tissue section. Bars represent mean  $\pm$  s.e.m. for three mice. Similar results were observed in three independent experiments. **b**, p62 immunoblot of brains of 5-day-old GFP-LC3 mice treated with the indicated peptide (20 mg kg<sup>-1</sup> i.p., 6 h). **c**, Survival curves of 5-day-old C57BL/6J mice infected with CHIKV (10<sup>6</sup> p.f.u. s.c.) and treated daily with peptide (15 mg kg<sup>-1</sup> i.p. beginning 1 day post-infection). **d–f**, Representative images of WNV envelope antigen and TdT-mediated dUTP nick end labelling (TUNEL) staining (**d**) (T, Tat alone), quantification of cell death in brain (**e**), and survival curves (**f**) for 5-day-old C57BL/6J mice infected with WNV (Egypt strain 101, 1 p.f.u. intracerebral (i.c.)) and treated daily with peptide (D-amino acid forms, 20 mg kg<sup>-1</sup> i.p. beginning 1 day post-infection). Images in **d** are from cerebral cortex day 6 post-infection. Similar results were observed in all regions of the brain for three mice per group. Scale bar, 20  $\mu\text{m}$ . Bars in **e** represent mean  $\pm$  s.e.m. TUNEL-positive cells per unit area of brain for three mice. **g**, Geometric mean + s.e.m. viral titres of WNV-infected mouse brains day 6 post-infection. Values represent combined data for 12–20 mice per treatment group from 10 to 12 litters. Data in **c** and **f** represent combined survival probabilities for three and four independent litters, respectively, in each group. Similar results were observed in each independent experiment. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; NS, not significant; *t*-test.

### Discussion

By mapping the domain of beclin 1 targeted by the HIV-1 pathogenicity protein, Nef, we identified a functionally important region of beclin 1, leading to the synthesis of a cell-permeable autophagy-inducing peptide and the discovery of a negative regulator of autophagy, GAPR-1. The autophagy-inducing activity of the peptide was associated with clearance of small polyglutamine expansion protein aggregates, reduced titres of several positive-stranded RNA viruses, decreased intracellular survival of the bacterium, *L. monocytogenes*, inhibition of HIV-1 replication in human macrophages, and a reduction in the mortality of neonatal mice infected with CHIKV and WNV. Because autophagy dysregulation may underlie the pathogenesis of metabolic, inflammatory and neoplastic diseases (in addition to



its role in infections and neurodegenerative disorders), specific autophagy-inducing agents such as the Tat–beclin 1 peptide may have potential for the prevention and treatment of a broad range of human diseases.

## METHODS SUMMARY

The active Tat–beclin 1 peptide sequence YGRKKRRQRRRGSTNVFNATFEIWHDEFGT consisted of 11 amino acids from the Tat protein transduction domain attached via a G2 linker to modified beclin 1 amino acids 267–284. Tat-scrambled (YGRKKRRQRRRGVGNDFFINHETTFATEW) was used as a control peptide. Autophagy was measured using immunoblot assays (p62 degradation, LC3 conversion), detection of GFP–LC3 puncta with fluorescence microscopy, radiolabelled long-lived protein degradation, and electron microscopy. For *in vitro* experiments to measure the effects of Tat–beclin 1 peptide on the clearance of CFP–htt103Q protein, viral replication, or bacterial replication, cells were treated with 10  $\mu$ M peptide for 2–4 h, except for the HIV-1 replication studies which used 0.5–5.0  $\mu$ M peptide for 10 days. For *in vivo* experiments, GFP–LC3 transgenic C57/BL6J mice (for autophagy measurements) or non-transgenic C57/BL6J mice (for infections with CHIKV or WNV) were treated i.p. with 15–20 mg kg<sup>−1</sup> of peptide.

**Full Methods** and any associated references are available in the online version of the paper.

**Received 2 March; accepted 17 December 2012.**

**Published online 30 January 2013.**

- Levine, B. & Kroemer, G. Autophagy in the pathogenesis of disease. *Cell* **132**, 27–42 (2008).
- Mizushima, N., Levine, B., Cuervo, A. M. & Klionsky, D. J. Autophagy fights disease through cellular self-digestion. *Nature* **451**, 1069–1075 (2008).
- Rubinshtein, D. C., Codogno, P. & Levine, B. Autophagy modulation as a potential therapeutic target for diverse diseases. *Nature Rev. Drug Discov.* **11**, 709–730 (2012).
- Kihara, A., Kabeya, Y., Ohsumi, Y. & Yoshimori, T. Beclin-phosphatidylinositol 3-kinase complex functions at the trans-Golgi network. *EMBO Rep.* **2**, 330–335 (2001).
- Kyei, G. B. *et al.* Autophagy pathway intersects with HIV-1 biosynthesis and regulates viral yields in macrophages. *J. Cell Biol.* **186**, 255–268 (2009).
- Furuya, N., Yu, J., Byfield, M., Pattingre, S. & Levine, B. The evolutionarily conserved domain of Beclin 1 is required for Vps34 binding, autophagy and tumor suppressor function. *Autophagy* **1**, 46–52 (2005).
- Huang, W. *et al.* Crystal structure and biochemical analyses reveal Beclin 1 as a novel membrane binding protein. *Cell Res.* **22**, 473–489 (2012).
- Liang, X. H. *et al.* Induction of autophagy and inhibition of tumorigenesis by *beclin 1*. *Nature* **402**, 672–676 (1999).
- Mizushima, N., Yoshimori, T. & Levine, B. Methods in mammalian autophagy research. *Cell* **140**, 313–326 (2010).
- van den Berg, A. & Dowdy, S. F. Protein transduction domain delivery of therapeutic macromolecules. *Curr. Opin. Biotechnol.* **22**, 888–893 (2011).
- Bayer, P. *et al.* Structural studies of HIV-1 Tat protein. *J. Mol. Biol.* **247**, 529–535 (1995).
- Lee, J. S. *et al.* FLIP-mediated autophagy regulation in cell death control. *Nature Cell Biol.* **11**, 1355–1362 (2009).
- Orvedahl, A. O. *et al.* Autophagy protects against Sindbis virus infection of the central nervous system. *Cell Host Microbe* **7**, 115–127 (2010).
- Eberle, H. B. *et al.* Identification and characterization of a novel human plant pathogenesis-related protein that localizes to lipid-enriched microdomains in the Golgi complex. *J. Cell Sci.* **115**, 827–838 (2002).
- Polson, H. E. *et al.* Mammalian Atg18 (WIPI2) localizes to omegasome-anchored phagophores and positively regulates LC3 lipidation. *Autophagy* **6**, 506–522 (2010).
- Harris, H. & Rubinshtein, D. C. Control of autophagy as a therapy for neurodegenerative disease. *Nature Rev. Neurol.* **8**, 108–117 (2012).
- Yamamoto, A., Cremona, M. L. & Rothman, J. E. Autophagy-mediated clearance of huntingtin aggregates triggered by the insulin-signaling pathway. *J. Cell Biol.* **172**, 719–731 (2006).
- Yoshikawa, Y. *et al.* *Listeria monocytogenes* ActA-mediated escape from autophagic recognition. *Nature Cell Biol.* **11**, 1233–1240 (2009).
- Zhao, Z. *et al.* Autophagosome-independent essential function for the autophagy protein Atg5 in cellular immunity to intracellular pathogens. *Cell Host Microbe* **4**, 458–469 (2008).
- Campbell, G. R. & Spector, S. A. Hormonally active vitamin D3 (1 $\alpha$ ,25-dihydroxycholecalciferol) triggers autophagy in human macrophages that inhibits HIV-1 infection. *J. Biol. Chem.* **286**, 18890–18902 (2011).
- Campbell, G. R. & Spector, S. A. Vitamin D inhibits human immunodeficiency virus type 1 and *Mycobacterium tuberculosis* infection in macrophages through the induction of autophagy. *PLoS Pathog.* **8**, e1002689 (2012).
- Mizushima, N., Yamamoto, A., Matsui, M., Yoshimori, T. & Ohsumi, Y. *In vivo* analysis of autophagy in response to nutrient starvation using transgenic mice expressing a fluorescent autophagosome marker. *Mol. Biol. Cell* **15**, 1101–1111 (2004).
- Fischer, P. M. The design, synthesis and application of stereochemical and directional peptide isomers: a critical review. *Curr. Protein Pept. Sci.* **4**, 339–356 (2003).
- Couderc, T. *et al.* A mouse model for Chikungunya: young age and inefficient type-I interferon signaling are risk factors for severe disease. *PLoS Pathog.* **4**, e29 (2008).
- Johnson, R. T. Acute encephalitis. *Clin. Infect. Dis.* **23**, 219–226 (1996).
- Griffin, D. E. Emergence and re-emergence of viral diseases of the central nervous system. *Prog. Neurobiol.* **91**, 95–101 (2010).
- Lauer, P., Chow, M. Y., Loessner, M. J., Portnoy, D. A. & Calendar, R. Construction, characterization, and use of two *Listeria monocytogenes* site-specific phage integration vectors. *J. Bacteriol.* **184**, 4177–4186 (2002).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank M. Diamond, J. L. Foster, M. Gale, N. Mizushima, D. Sabatini, M. Shiloh and T. Yoshimori for supplying critical reagents; and H. Ball, A. Bugde and E.-L. Eskelinen for assistance with peptide synthesis, infrared imaging and EM interpretation, respectively. This work was supported by NIH grants U54AI057156 (B.L.), K08 AI099150 (R.S.), RO1 NS077874 (S.A.S.), RO1 GM094575 (N.V.G.), RO1 GM066099 (O.L.), RO1 GM079656 (O.L.), RO1 NS063973 (A.Y.), RO1 NS050199 (A.Y.), U54AI057160 (H.W.V., D.J.L.), RO1 DK083756 (R.X.), RO1 DK086502 (R.X.), and T32 GM008297 (C.H.); NSF CCF-0905536 (O.L.); an NWO-ALW Open Program Grant 817.02.023 (J.B.H.); Cancer Research UK (S.A.T.); and a Welch Foundation Award I-15-5 (N.V.G.).

**Author Contributions** B.L., S.S.-K. and O.L. generated the original hypothesis. S.S.-K., R.S., M.L., G.R.C., Z.Z., Q.S., K.P., D.M., C.H., R.E., D.K. and D.V.K. performed experiments. L.K., A.D.W., R.X., O.L. and N.V.G. performed bioinformatics analyses. B.L., S.S.-K., R.S., M.L., L.K., H.W.V., J.B.H., S.A.T., R.X., D.J.L., A.Y., O.L., N.V.G., S.A.S. and D.V.K. provided intellectual contributions throughout the project. B.L., S.S.-K. and R.S. took primary responsibility for writing the manuscript. All authors edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.L. ([beth.levine@utsouthwestern.edu](mailto:beth.levine@utsouthwestern.edu)).

## METHODS

**Cell culture.** COS-7, HeLa, A549 and MCF7 cells were cultured in DMEM containing 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. HBEC30-KT cells were cultured in keratinocyte-SFM with 50 mg ml<sup>-1</sup> bovine pituitary extract, 5 ng ml<sup>-1</sup> recombinant epidermal growth factor (Gibco) and 1% penicillin/streptomycin<sup>28</sup>. THP-1 and HCC827 cells were cultured in RPMI 1640 containing 10% FBS and 1% penicillin/streptomycin. For monocytic differentiation, THP-1 cells were incubated with 10 nM PMA (Sigma) for 72 h. MEFs were cultured in DMEM containing 15% FBS, 1% penicillin/streptomycin, 118  $\mu$ M  $\beta$ -mercaptoethanol, and 1 $\times$  MEM non-essential amino acids. HeLa/GFP-LC3 cells and HeLa-htt103Q cells have been described previously<sup>13,17</sup>. Murine BMDMs were cultured in DMEM containing 20% heat-inactivated FBS, 20% L-cell supernatant, and 1% penicillin/streptomycin. *Atg5<sup>fllox/fllox</sup>* and *Atg5<sup>fllox/fllox</sup>-LysM-Cre* BMDMs have been described<sup>19</sup>. Primary human MDMs were prepared from whole blood and cultured in RPMI 1640 containing 10% charcoal/dextran-treated, heat-inactivated FBS (Gemini Bio-Products) and 10 ng ml<sup>-1</sup> macrophage CSF (R&D Systems) for 9 days as described<sup>21</sup>. Venous blood was obtained from HIV-1 seronegative subjects using a protocol approved by the Human Research Protections Program of the University of California, San Diego (Project 08-1613) in accordance with the Code of Federal Regulations on the Protection of Human Subjects (45 CFR part 46 and 21 CFR part 50 and 56). Written informed consent was obtained from blood donors before their participation.

**Plasmid construction and transfection.** pCDNA3.1 Myc-Vps34 has been described previously<sup>6</sup>. *BECLN1* cDNA was subcloned from pCR3.1-beclin 1 (ref. 29) into pBICEP-CMV2 (Sigma). Truncated constructs of Flag-beclin 1 were subcloned as PCR products containing each domain into pBICEP-CMV2. Deletion constructs of Flag-beclin 1 and the Flag-beclin 1 (H275E/S279D/Q281E) mutant were generated by PCR mutagenesis and subcloned into pBICEP-CMV2. The HIV-1 SF2 Nef plasmid, pCDNA-Nef<sup>30</sup>, was provided by J. L. Foster. The Nef protein was fused with a 2 $\times$  HA sequence at the C terminus by PCR and subcloned into pCDNA3.1. *GAPR-1* cDNA from pCMV6-XL4 (Origene) was fused with a Myc sequence at the C terminus by PCR and subcloned into pCDNA3.1. Cells were transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Immunoprecipitation and fluorescent microscopy analysis were performed 20–27 h post-transfection. GFP-LC3<sup>31</sup> and mRFP-GFP-LC3<sup>32</sup> plasmids were provided by T. Yoshimori.

**Antibodies and reagents.** For immunoblot analyses, rat monoclonal anti-HA (1:500 dilution, Roche), mouse monoclonal anti-Flag M2 (1:500 dilution, Sigma), mouse monoclonal anti-Myc (1:500 dilution, Santa Cruz), rabbit anti-LC3 (1:300 dilution, Novus Biologicals), mouse monoclonal anti-p62 (1:500 dilution, BD Biosciences), guinea-pig anti-p62 (1:500 dilution, Progen), mouse monoclonal anti-actin (1:2,000 dilution, Santa Cruz), rabbit anti-beclin 1 (1:1,000 dilution, Santa Cruz), rabbit anti-ATG7 (1:1,000 dilution, (Sigma), anti-ATG5 (1:500 dilution, Novus Biologicals), anti-GFP (1:500 dilution, Invitrogen) and rabbit anti-GAPR-1 (1:10,000 dilution)<sup>14</sup> were used. 3-MA (Sigma) was dissolved in H<sub>2</sub>O and used at a 10  $\mu$ M concentration. Torin1 (gift from D. Sabatini)<sup>33</sup>, rapamycin (Sigma) and bafilomycin A1 (Sigma) were dissolved in DMSO and used at concentrations of 1  $\mu$ M, 500 nM and 100 nM, respectively.

**Immunoblot analyses.** Cells were rinsed with Dulbecco's phosphate buffered saline (PBS(-), (Sigma)) and lysed in lysis buffer (20 mM HEPES, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, protease inhibitor cocktail (Roche)) on ice for 1 h. Lysates were centrifuged at 16,100g for 10 min at 4 °C, and resolved by SDS-PAGE and transferred to PVDF membranes. Membranes were blocked with 5% non-fat dry milk (NFDm) in PBST (PBS 0.05% Tween-20), incubated overnight at 4 °C with primary antibodies in PBST containing 5% NFDm, washed with PBST, and incubated for 1 h at room temperature with HRP-conjugated secondary antibodies diluted in PBST containing 5% NFDm. Immunoreactive bands were visualized with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) or Amersham ECL Plus Western Blotting Detection System (GE Healthcare).

For detection of beclin 1 and GAPR-1 in Golgi fractions of HeLa cells and HeLa/GAPR-1-Myc cells, Golgi-enriched fractions were isolated as described<sup>34,35</sup> from cells treated with peptide (20  $\mu$ M, 1 h), then trypsinized and homogenized with a Balch homogenizer (gap size 9  $\mu$ m). Post-nuclear supernatant (PNS) was obtained after centrifugation of the cell homogenate (1,700 r.p.m., 10 min, 4 °C). The PNS was fractionated by sucrose density gradient ultracentrifugation (SV40 rotor at 100,000g, 2 h 40 min) and approximately 1 ml of a Golgi-enriched fraction was collected at the 35–29% sucrose interface. Membranes were pelleted by ultracentrifugation (100,000g, 30 min) after the addition of three volumes of PBS to one volume of Golgi-enriched fraction.

**Immunoprecipitation.** Cell lysates were centrifuged at 16,100g for 10 min at 4 °C and the supernatants were subjected to immunoprecipitation with an anti-HA

antibody (1:100 dilution, Roche) and protein G plus-agarose (Santa Cruz) or with anti-Myc agarose (1:20 dilution, Santa Cruz) and subjected to immunoblot analyses.

**Cell viability assays.** Cytotoxicity was measured in HeLa cells using an LDH cytotoxicity kit (Cayman Chemical). Cytotoxicity in primary human MDMs was measured using WST-1 (Roche). Cell survival in primary murine BMDMs was measured by trypan blue exclusion or a CellTiter-Glo Luminescent Cell Viability assay kit (Promega).

**siRNA and shRNA suppression of gene expression.** siRNAs were purchased from Thermo Scientific Dharmacon. siRNAs specific to *BECLN1* (5'-GCACUC AAGUUC AUGCUUA-3'), *ATG7* (5'-GGGUU AUUACUACAAUGGUGUU-3'), *C9orf19* (*GAPR-1*) (no. 2, 5'-GCAAGAACCUCAACCGGGA-3' and no. 17, 5'-GUUACCUAGACCACGAUUA-3') and a non-silencing control siRNA (NC-2) were transfected using Lipofectamine 2000. Immunoblot and fluorescent microscopic analyses were performed 48–72 h post-transfection. MISSION short-hairpin (shRNA) lentiviral particles were from Sigma: *ATG5* (SHCLNV-NM\_004849/TRCN0000150940), or scrambled non-targeting negative control (Scr, SHC002V).

**Fluorescence microscopy and image analysis.** For measurement of autophagy in HeLa/GFP-LC3 cells or MCF7 cells, cells were fixed with 2% paraformaldehyde (PFA) in PBS, and GFP-LC3-positive dots per cell was quantified as described<sup>29</sup>. To measure autophagic flux, HeLa cells were co-transfected with an mRFP-GFP-LC3 plasmid<sup>32</sup> and siRNA, treated with peptide (25  $\mu$ M, 3 h) 72 h post-transfection, and then fixed as above. GFP-positive autophagosomes and RFP-positive autolysosomes per cell were quantified. For the htt103Q aggregate assay, CFP-positive aggregates <1  $\mu$ m were counted.

For immunofluorescence (IFA) analysis of cultured cells, HeLa cells were fixed as above, permeabilized in 0.5% Triton X-100/PBS, blocked with 0.5% BSA in PBS, stained with anti-WIP1<sup>15</sup> (mouse, 1:500), anti-beclin 1 (rabbit, 1:250, Santa Cruz), anti-GM130 (mouse, 1:250, Novus Biologicals), and labelled with goat-anti-rabbit AlexaFluor 647 and goat-anti-mouse AlexaFluor 488 secondary antibodies (1:500) or stained with anti-streptavidin-AlexaFluor 488 (1:500) (after treatment with Tat-biotin-conjugated peptides). For the beclin 1 primary antibody, specificity was confirmed by siRNA knockdown. Z-stacks were acquired with a Zeiss AxioImager M2 microscope equipped with a Photometrics CoolSnap HQ2 camera using a Zeiss PLAN APO 20X/0.8 NA wide-field objective with the same acquisition times for samples stained with identical primary antibody pairs, then deconvolved with AutoDeBlur (Biotplane). Imaris version 7.4.0 (Bitplane) was used for analysis. Thresholding for background secondary antibody staining was performed in all experiments. Representative images were chosen after examining 60–120 cells from triplicate samples for each condition. Nuclear staining (besides DAPI) was masked using the DAPI channel and setting non-DAPI signal within the nuclear surface to zero. Quantification of WIP12-positive dots per cell was performed by an observer blinded to experimental condition.

For IFA of mouse tissues, 5-day-old GFP-LC3 transgenic mice infected with CHIKV were treated with peptide (20 mg kg<sup>-1</sup>, 6 h). Muscle sections were prepared as described in the section below, stained with an anti-CHIKV E2 antibody<sup>13,36</sup>, and images were acquired as described for cultured cells above. For infrared (IR) dye images, muscle tissue was collected from peptide-treated GFP-LC3 mice (25 mg kg<sup>-1</sup> i.p., 2 h), and sections were stained with streptavidin-conjugated IRDye 800CW (1:500 dilution; Licor). Images were acquired with a Zeiss AxioObserver Z1 microscope equipped with a Hamamatsu Orca-ER camera using a Zeiss PLAN APO 63 $\times$ /1.4 NA oil objective and analysed by an observer blinded to experimental condition.

**Electron microscopy.** HeLa cells were fixed overnight with 2.5% glutaraldehyde in 0.1 M (pH 7.4) cacodylate buffer and samples were prepared and analysed as described<sup>8</sup>.

**Immunohistochemistry.** Mouse brain sections were fixed in 4% PFA. Paraffin-embedded sagittal sections were TUNEL-stained to detect apoptotic nuclei (Apoptag peroxidase *in situ* Apoptosis Detection Kit; Chemicon International). Immunoperoxidase staining was performed to detect WNV antigen (rat polyclonal anti-WNV antibody; 1:100 dilution; provided by M. Diamond).

**Peptide design.** An evolutionary trace (ET) analysis<sup>37</sup> revealed regions within the 267–299 domain of beclin 1 likely to be of evolutionary importance. To enhance the solubility of the Tat PTD (amino acids 47–57)<sup>10</sup> fused to the beclin 1 267–284 fragment, ET analysis of beclin 1 further identified residue positions unlikely to be functionally important occupied by charged residues in beclin 1 orthologues. These cognate residues were then substituted into the peptide (H275E, S79D and Q281E) to increase its solubility without disrupting its mimicry of beclin 1 function.

**Peptide synthesis.** L-amino acid peptides were synthesized by the University of Texas Southwestern Medical Center (UTSW) Protein Chemistry Technology Core and purified to >95% by HPLC (confirmed by mass spectrometry). The Tat-beclin 1 peptide sequence, YGRKKRRQRRRGGTNVFNATFEIWHDG EFGT, consisted of 11 amino acids from the Tat PTD at the N terminus, a GG



linker to increase flexibility, and at the C terminus, 18 amino acids derived from beclin 1 267–284 containing three substitutions, including H275E, S279D, Q281E. Wild-type Tat–beclin 1 peptide consisted of the Tat PTD, a GG linker, and 18 amino acids derived from the natural beclin 1 sequence (267–284) (YGRKKRRQRRRGSTNVNATFHIWHSQGFGT). Control peptide, Tat-scrambled, consisted of the Tat protein transduction domain, a GG linker, and a scrambled version of the C-terminal 18 amino acids from Tat–beclin 1 (YGRKKRRQRRRGVGNDFFINHETTFATEW). For experiments comparing Tat–beclin 1 and Tat-scrambled, peptides were dissolved in PBS(–), whereas in the experiment comparing Tat–beclin 1 and wild-type Tat–beclin 1, peptides were dissolved in H<sub>2</sub>O. Peptides were stored at –80 °C. For peptide treatment, cells were washed with PBS(–) and treated with peptides (10–50 µM, 1–4 h) dissolved in OPTI-MEM (Gibco) acidified with 0.15% (v/v) 6N HCl. For treatment of primary human MDMs, cells were washed with PBS and pre-treated with peptides (0.5–5 µM, 24 h) in 500 µl macrophage-SFM (serum-free media) (Gibco) before infection with HIV-1.

D-amino acid peptides were synthesized at the HHMI Mass Spectrometry Laboratory at UC-Berkeley. The retro-inverso Tat–beclin 1 D-amino acid sequence was RRRQRRKKRGYGGTGFEGDHWIEFTANFVNT. Peptides were synthesized by solid-phase methodology on Wang resin of 0.44 meq g<sup>–1</sup> substitution using an ABI 431A synthesizer. Appropriate N-FMOC amino acid derivatives were coupled via dicyclohexylcarbodiimide activation in dichloromethane/N-methylpyrrolidone using user-devised extended activation, coupling and piperidine deprotection cycles. Dried resin-peptide was deprotected in reagent K 4 h at room temperature. Peptide was extracted with warm acetonitrile/water, lyophilized, and purified by RPLC. Crude peptide purity was roughly 75%; after purification, 95%. D-amino acid peptides were dissolved in H<sub>2</sub>O and stored at –80 °C until use. Peptide purity was assessed by FTICR mass spectrometry.

**Circular dichroism spectroscopy.** CD spectra of the Tat–beclin 1 and Tat-scrambled peptides were recorded on an AVIV 62DS spectropolarimeter.

**Model of beclin 1 peptide.** The beclin 1 peptide model was adopted from the corresponding elements of the ECD structure<sup>7</sup> (Protein Data Bank accession 4DDP). The first two residues after the diglycine linker are disordered in the ECD structure and were omitted. The side chains of the residues mutated to increase solubility were modelled using the mutagenesis wizard of PyMOL, keeping rotamers with no steric clashes.

**Measurement of long-lived protein degradation.** COS-7 cells cultured in 6-well plates were incubated in 10% FBS/DMEM containing L-<sup>3</sup>H leucine (1 µCi ml<sup>–1</sup>) (Perkin Elmer) and 65 µM of unlabelled leucine for 3 days. Cells were washed with 10% FBS-DMEM three times and incubated with 10% FBS/DMEM containing unlabelled leucine (2 µM) for 1 day. After washing three times with normal medium, cells were treated with indicated peptide or EBSS for 2 h with or without 3-MA. The medium was precipitated with 10% TCA and TCA-soluble radioactivity was measured. Cells were fixed by 10% TCA, lysed in 0.2N NaOH and total cell radioactivity was measured. L-<sup>3</sup>H-leucine release was estimated as a percentage of the radioactivity in the TCA-soluble material relative to the total cell radioactivity.

**Protein purification and mass spectrometry.** Tat-scrambled and Tat–beclin 1 peptides were conjugated at their N termini to biotin via a PEG(3) linkage by the UTSW Protein Chemistry Technology Core. Approximately 8 × 10<sup>6</sup> HeLa cells were treated with biotinylated peptides (30 µM, 3 h), lysed in lysis buffer (20 mM HEPES, 137 mM NaCl, 1 mM EDTA, 0.2% Triton X-100, 10% glycerol and protease inhibitor cocktail) and homogenized using a 25G syringe. Lysates were mixed with M-280 streptavidin Dynabeads (Invitrogen) overnight and 20 µl of laemmli sample buffer was added. After separation by SDS–PAGE, bands were visualized by GelCode blue stain reagent (Thermo Scientific), excised and analysed by LC-MS/MS at the UTSW Protein Chemistry Technology Core. All MS/MS spectra were searched against the protein sequence database at the NCBI using Mascot software.

**Filter trap assay.** A filter trap assay was performed to measure amounts of soluble htt103Q protein, small htt103Q aggregates (P2 supernatant) and large htt103Q aggregates (P1 supernatant) in HeLa-htt103Q cells as described<sup>17,38</sup>.

**Virus strains and infection.** SINV strain SV1A (ATCC) has been described previously<sup>39</sup>. WNV strain TX02<sup>40</sup> was provided by M. Gale and WNV strain Egypt 101<sup>41</sup> was provided by M. Diamond. CHIKV strain CHIKV-21 was propagated as described<sup>42</sup>. All *in vitro* virus infections were performed at a multiplicity of infection (MOI) of 0.1 in reduced serum (1% heat-inactivated FBS) containing media for 1 h. For titration of viruses in mouse tissues, 10% homogenates of freeze-thawed tissue were used. Viral stocks, cell culture supernatants, and tissue homogenates were titred by plaque assay titration using BHK-21 cells for SINV and CHIKV and Vero cells for WNV.

For HIV-1 experiments, 5 × 10<sup>5</sup> MDMs were treated with peptide at the indicated dose for 24 h, infected with HIV-1<sub>Ba-L</sub> (provided by S. Gartner and R. Gallo)<sup>43</sup> at an MOI of 0.01 for 3 h, washed 3× with PBS, and then incubated with macrophage-SFM containing the peptide for 10 days. At days 3, 5, 7 and 10, 250 µl of cell supernatant was collected and fresh 250 µl macrophage-SFM containing the peptides was added. Extracellular release of HIV-1 p24 antigen into the supernatants was measured using the Alliance HIV-1 p24 antigen enzyme-linked immunosorbent assay kit (ELISA; PerkinElmer).

**Listeria infection.** Primary BMDMs were collected from C57/BL6J mice and cultured for 6 days. On day 7, BMDMs were plated in DMEM containing 20% FBS and 20% L-cell supernatant and no antibiotics, and infected 24 h later with the *Listeria*  $\Delta$ actA mutant strain DPL-4029<sup>27</sup> (provided by M. Shiloh) at an MOI of 5 c.f.u. per cell for 30 min at 37 °C and 5% CO<sub>2</sub>, then gently washed three times with PBS and treated with either 10 µM Tat-scrambled or Tat–beclin 1 peptide for 2 h. After 2 h, cells were washed three times with PBS to remove extracellular bacteria and peptide and the medium was replaced with non-antibiotic-containing medium. BMDMs were lysed in 0.3% Triton X-100 at serial time points and serial dilutions were plated on LB agar overnight at 37 °C for determination of c.f.u.

**Animal experiments.** To measure autophagy in mouse tissues, 6-week-old GFP–LC3 transgenic mice<sup>22</sup> were injected i.p. with Tat-scrambled (L-amino acid), Tat–beclin 1 (L-amino acid) and Tat–beclin 1 (D-amino acid) peptide at 20 mg kg<sup>–1</sup> (5.3 µmol kg<sup>–1</sup>) or with Tat (D-amino acid) peptide at 5.3 µmol kg<sup>–1</sup>. After 6 h, mice were killed and fixed by perfusion with 4% PFA in PBS. Tissues were fixed in 4% PFA overnight, 15% sucrose for 4 h, and 30% sucrose overnight before frozen sections were prepared and used for fluorescence microscopy analysis as described<sup>44</sup>. For immunoblot analysis, 5-day-old GFP–LC3 mice were injected i.p. with Tat-scrambled (L-amino acid), Tat–beclin 1 (L-amino acid) and Tat–beclin 1 (D-amino acid) peptide at 20 mg kg<sup>–1</sup> (5.3 µmol kg<sup>–1</sup>) or with Tat (D-amino acid) peptide at 5.3 µmol kg<sup>–1</sup>. Six hours later, mice were killed and frozen brain tissue homogenates were used. For virus infections, CHIKV and WNV Egypt strain were diluted in HBSS and used to infect 5-day-old C57BL/6J mice with 10<sup>6</sup> p.f.u. s.c or 1 p.f.u. i.c., respectively. For assessment of peptide toxicity, 6-day-old C57BL/6J mice were injected i.p. with Tat-scrambled (L-amino acid) or Tat–beclin 1 (L-amino acid) at 15 mg kg<sup>–1</sup> and with Tat (D-amino acid) at 5.3 mmol kg<sup>–1</sup> or Tat–beclin 1 (D-amino acid) at 20 mg kg<sup>–1</sup> daily for 2 weeks. Body weight, neurological status and general clinical status was monitored daily, and blood, liver, spleen, heart and brain samples were collected upon death after 2 weeks of peptide treatment. For serum biochemistry and haematological analyses in adult mice, 3-month-old C57BL/6J mice were injected i.p. with Tat-scrambled (L-amino acid) or Tat–beclin 1 (L-amino acid) at 20 mg kg<sup>–1</sup> daily for 2 weeks before blood collection. Animal experiments were approved by the UTSW Institutional Animal Care Use Committee and performed in accordance with institutional guidelines.

**Retroviral expression system.** *GAPR-1* (also known as *GLIPR2*) cDNA (Origene) was subcloned into pBabe-puro (Addgene) by PCR. To generate pBabe-GAPR-1-Myc, *GAPR-1* cDNA was fused with a Myc sequence at the C terminus by PCR and subcloned into pBabe-puro. To produce retroviruses, 4 µg of pBabe-GAPR-1 was transfected into 293T cells with 3.6 µg of pUMVC (Addgene) and 0.4 µg of pCMV-VSV-G (Addgene). At 2 and 3 days after transfection, the viral supernatants were collected and filtered through a 0.45-µm membrane filter and the viral supernatants were used to infect HeLa cells in the presence of 6 µg ml<sup>–1</sup> of polybrene. At 24 h post-infection, fresh viral supernatants with polybrene were added to the cells for a second round of infection. After an additional 24 h, cells were cultured with 10% FBS/DMEM with 4 µg ml<sup>–1</sup> puromycin.

28. Sato, M. *et al.* Multiple oncogenic changes (K-RAS(V12), p53 knockdown, mutant EGFRs, p16 bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells. *Cancer Res.* **66**, 2116–2128 (2006).
29. Pattingre, S. *et al.* Bcl-2 antiapoptotic proteins inhibit Beclin 1-dependent autophagy. *Cell* **122**, 927–939 (2005).
30. O'Neill, E. *et al.* Dynamic evolution of the human immunodeficiency virus type 1 pathogenic factor, Nef. *J. Virol.* **80**, 1311–1320 (2006).
31. Kabeya, Y. *et al.* LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. *EMBO J.* **19**, 5720–5728 (2000).
32. Kimura, S., Noda, T. & Yoshimori, T. Dissection of the autophagosome maturation process by a novel reporter protein, tandem fluorescent-tagged LC3. *Autophagy* **3**, 452–460 (2007).
33. Thoreen, C. C. *et al.* An ATP-competitive mammalian target of rapamycin inhibitor reveals rapamycin-resistant functions of mTORC1. *J. Biol. Chem.* **284**, 8023–8032 (2009).
34. Balch, W. E., Dunphy, W. G., Braell, W. A. & Rothman, J. E. Reconstitution of the transport of protein between successive compartments of the Golgi

- measured by the coupled incorporation of N-acetylglucosamine. *Cell* **39**, 405–416 (1984).
35. Vogels, M. W. *et al.* Quantitative proteomic identification of host factors involved in the *Salmonella typhimurium* infection cycle. *Proteomics* **11**, 4477–4491 (2011).
  36. Bréhin, A. C. *et al.* Production and characterization of mouse monoclonal antibodies reactive to Chikungunya envelope E2 glycoprotein. *Virology* **371**, 185–195 (2008).
  37. Mihalek, I., Res, I. & Lichtarge, O. A family of evolution-entropy hybrid methods for ranking protein residues by importance. *J. Mol. Biol.* **336**, 1265–1282 (2004).
  38. Bailey, C. K., Andriola, I. F., Kampinga, H. H. & Merry, D. E. Molecular chaperones enhance the degradation of expanded polyglutamine repeat androgen receptor in a cellular model of spinal and bulbar muscular atrophy. *Hum. Mol. Genet.* **11**, 515–523 (2002).
  39. Taylor, R. M., Hurlbut, H. S., Work, T. H., Kingston, J. R. & Frothingham, T. E. Sindbis virus: a newly recognized arthropod-transmitted virus. *Am. J. Trop. Med. Hyg.* **4**, 844–862 (1955).
  40. Keller, B. C. *et al.* Resistance to alpha/beta interferon is a determinant of West Nile virus replication fitness and virulence. *J. Virol.* **80**, 9424–9434 (2006).
  41. Weiner, L. P., Cole, G. A. & Nathanson, N. Experimental encephalitis following peripheral inoculation of West Nile virus in mice of different ages. *J. Hyg. (Lond.)* **68**, 435–446 (1970).
  42. Schuffenecker, I. *et al.* Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. *PLoS Med.* **3**, e263 (2006).
  43. Gartner, S. *et al.* The role of mononuclear phagocytes in HTLV-III/LAV infection. *Science* **233**, 215–219 (1986).
  44. Qu, X. *et al.* Promotion of tumorigenesis by heterozygous disruption of the *beclin 1* autophagy gene. *J. Clin. Invest.* **112**, 1809–1820 (2003).



# The structure of the asteroid 4 Vesta as revealed by models of planet-scale collisions

M. Jutzi<sup>1</sup>, E. Asphaug<sup>2</sup>, P. Gillet<sup>3</sup>, J.-A. Barrat<sup>4</sup> & W. Benz<sup>1</sup>

Asteroid 4 Vesta seems to be a major intact protoplanet, with a surface composition similar to that of the HED (howardite-eucrite-diogenite) meteorites<sup>1–4</sup>. The southern hemisphere is dominated by a giant impact scar<sup>5</sup>, but previous impact models<sup>6–8</sup> have failed to reproduce the observed topography. The recent discovery that Vesta's southern hemisphere is dominated by two overlapping basins<sup>9</sup> provides an opportunity to model Vesta's topography more accurately. Here we report three-dimensional simulations of Vesta's global evolution under two overlapping planet-scale collisions. We closely reproduce its observed shape, and provide maps of impact excavation and ejecta deposition. Spiral patterns observed in the younger basin Rheasilvia<sup>9</sup>, about one billion years old<sup>10</sup>, are attributed to Coriolis forces during crater collapse. Surface materials exposed in the north come from a depth of about 20 kilometres, according to our models, whereas materials exposed inside the southern double-excavation come from depths of about 60–100 kilometres. If Vesta began as a layered, completely differentiated protoplanet, then our model predicts large areas of pure diogenites and olivine-rich rocks. These are not seen<sup>11–13</sup>, possibly implying that the outer 100 kilometres or so of Vesta is composed mainly of a basaltic crust (eucrites) with ultramafic intrusions (diogenites).

Detailed views of Vesta reveal an interesting and unexpected geology<sup>14</sup>. The south polar depression is deeper and larger than estimated from previous Hubble Space Telescope<sup>5</sup> observations, and consists of two overlapping giant craters<sup>9</sup>. Howardites—mixtures of eucrites (basalt-like rocks) and diogenites (pyroxene-rich rocks)—are the most representative rock type observed on Vesta's surface. Multispectral images reveal a hemispherical-scale dichotomy<sup>12,13</sup>, as well as local variations in the concentration of eucrite and diogenite material.

The two overlapping southern basins, Veneneia followed by the younger Rheasilvia, probe Vesta's internal structure to depth of the order of 100 km, according to impact scaling laws<sup>15</sup>. In a study<sup>7</sup> published before the arrival of NASA's Dawn mission at Vesta, we modelled the formation of the Rheasilvia basin and reported that Vesta's fast rotation significantly deforms the pattern of the ejecta fallback. Ejecta from such collisions are massive and globally distributed<sup>7</sup>, in part due to the asteroid's rapid rotation (rotation period  $P_{\text{Vesta}} = 5.3$  h), with exhumed and deposited materials dominating the surface mineralogy. Understanding the provenance and specific distribution of ejecta is therefore key to understanding the surface mineralogy of Vesta.

In our previous work<sup>7</sup>, the pre-impact rotation axis was chosen arbitrarily. This is remedied by Dawn's discovery of an earlier huge basin (Veneneia); we now assume that the spin axis would have aligned to Veneneia before the formation of Rheasilvia. We study the successive formation of the Veneneia and Rheasilvia basins, using a three-dimensional smooth particle hydrodynamics (SPH) impact code<sup>7,16,17</sup>. Our modelling begins with a spherical, monolithic, differentiated non-rotating asteroid. The formation of Veneneia leads to a fully shattered Vesta, with a basin of roughly 400 km diameter and a small central peak. We then use a fully damaged rheology, and a spin along Veneneia's axis,

for modelling the second basin, Rheasilvia. Including pre-impact rotation and taking into account the presence of the underlying Veneneia basin distinguishes our simulations from two-dimensional axisymmetric modelling approaches (see, for example, refs 6, 8, 18).

The model results are the outcome of two successive major-scale basin forming collisions, Veneneia followed by Rheasilvia, referenced to an original spherically symmetric differentiated protoplanet (Fig. 1). The diameter of the modelled Rheasilvia basin, though difficult to map (in the model as on Vesta), is consistent to  $\sim 10\%$  with the new observations by Dawn<sup>19</sup>. Underlying the second structure is the Veneneia basin from the first simulation; the older basin is partly covered by ejecta from Rheasilvia but is still partly visible in Fig. 1. Successive ejection of the same terrain digs deeper into Vesta.

Given the large parameter space of possible impact conditions, not every detail of Vesta's shape is reproduced in the simulations. For instance, the observations by Dawn show higher elevations in the southwestern quadrant, whereas in the simulations we find higher elevations in the opposite quadrant. These differences might be explained by different impact angles, a more complex, non-spherical initial shape of the target, or a different rotation axis. For example, whereas the Rheasilvia impact azimuth relative to the spin axis is not arbitrarily chosen, the projectile's impact angle is so chosen.

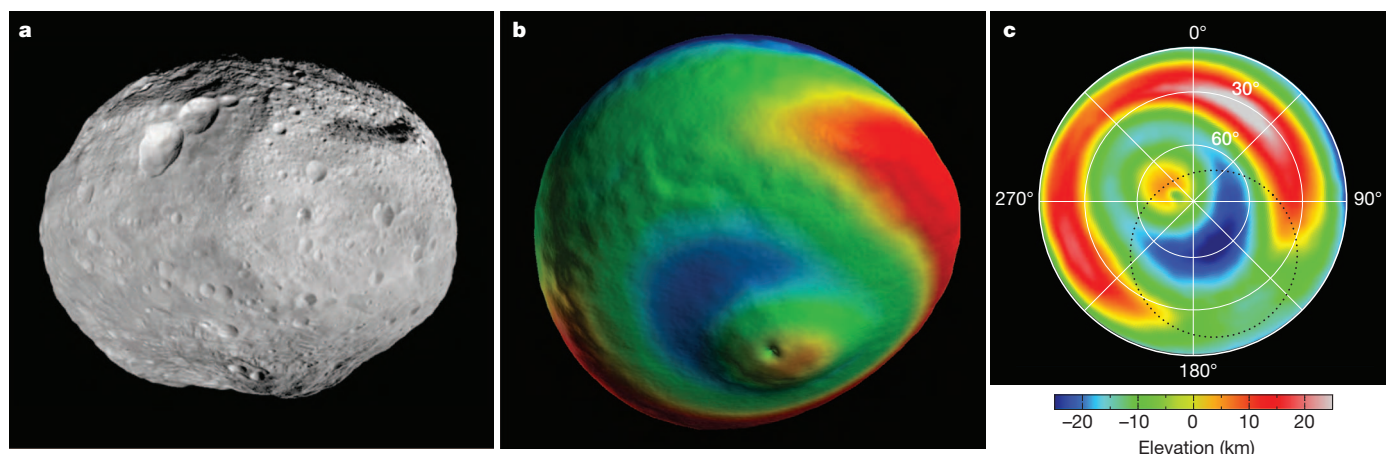
On a fast-spinning body, the Coriolis acceleration can become important for large-scale motions. The relative importance of the Coriolis acceleration is given by the Rossby number  $Ro = U/(Lf)$ , where  $U$  is the velocity,  $L$  the length scale and  $f = 2\omega_0 \sin(\varphi)$ ; here  $\omega_0$  is the rotation frequency, and  $\varphi$  the latitude. For small  $Ro$ , rotational effects are important. In the case of the gravitational collapse of Rheasilvia, typical velocities are  $\sim 50 \text{ m s}^{-1}$ , the length scale is  $\sim 200 \text{ km}$  and  $\omega_0 \approx 3.27 \times 10^{-4} \text{ s}^{-1}$ . For  $\varphi = 40^\circ$ ,  $Ro \approx 0.5$ , which indicates that Coriolis forces were important. This is confirmed by our simulations, which show that Vesta's rotation leads to a clockwise spiral pattern in the velocity field (see Fig. 2) that is similar (in both location and orientation) to the well-organized (clockwise) spiral fracture patterns observed inside the Rheasilvia crater<sup>9</sup>. This indicates that they probably formed during crater collapse under the influence of Coriolis accelerations. Other asymmetric features observed on the surface of Vesta<sup>9,13</sup> can be explained by this effect (Supplementary Fig. 1). Ejecta are also affected by the target rotation<sup>7</sup>, resulting in asymmetries in their distribution on Vesta's surface.

If the crust of Vesta formed from the cooling of a magma ocean<sup>20</sup>, it should be layered, with a 25–40-km-thick eucritic carapace overlying a 13–40-km-thick layer of diogenites<sup>21</sup>. At greater depth, olivine-rich lithologies ('mantle') and a metal core are expected<sup>21,22</sup>. This simple magma ocean scenario is still debated, and more complex models are possible. Diogenites could have crystallized in plutons (see, for example, ref. 23) intruding the eucritic crust (see, for example, refs 24, 25) rather than forming a uniform layer.

Interior models can be connected directly to the Dawn mineralogical surface data through our impact modelling, because it gives

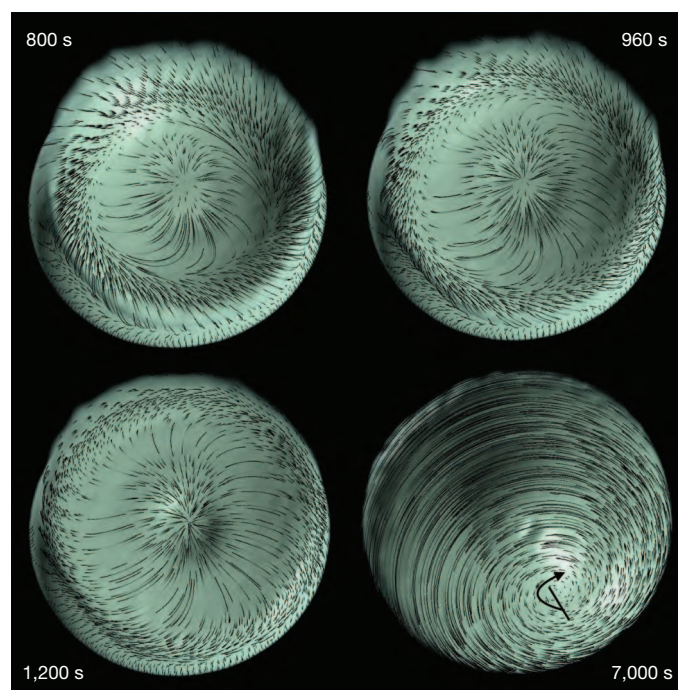
<sup>1</sup>Physics Institute, Space Research and Planetary Sciences, Center for Space and Habitability, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland. <sup>2</sup>School of Earth and Space Exploration, Arizona State University, PO Box 876004, Tempe, Arizona 85287, USA. <sup>3</sup>Institute of Condensed Matter Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL) Station 1, 1015 Lausanne, Switzerland.

<sup>4</sup>Université de Bretagne Occidentale, Institut Universitaire Européen de la Mer, CNRS UMR 6538, Place Nicolas Copernic, 29280 Plouzané, France.



**Figure 1 | SPH simulation of the formation of the two giant impact features in Vesta's southern hemisphere.** **a**, Asteroid Vesta as seen by Dawn; **b**, final result of the simulation; and **c**, Lambert azimuthal projection (equal area) of the southern hemisphere. Colours in **b** and **c** indicate the elevation (in kilometres; see key in **c**) with respect to a reference ellipsoid of  $280 \times 280 \times 230$  km. Note that the simulation data (size of the body) was scaled by a factor of 1.04. The dotted circle in **c** corresponds to the older Veneneia basin, which is partly overlapped by the younger and larger Rheasilvia giant impact feature. Overall, our model results are in good agreement with the observations by Dawn<sup>19</sup>. To model the impact events, we use an SPH impact code specially written to model geologic materials<sup>7,16,17</sup>. We include a tensile fracture model<sup>16</sup> in combination with a standard Drucker–Prager yield criterion for rocky materials. Damaged material is modelled using the Coulomb dry-friction law<sup>28</sup>. The block-model

approximation of acoustic fluidization<sup>29</sup> is used, as it gives a better match to central peak formation. Self-gravity is computed using a grid-based solver. The initial target is a non-rotating sphere of diameter  $d = 550$  km. To simulate the formation of Veneneia, a projectile of diameter  $d = 64$  km, impact velocity of  $5.4 \text{ km s}^{-1}$  and an impact angle of  $90^\circ$  (that is, head-on impact) is used. The outcome of this simulation is then used as initial condition for a second run where we study the formation of Rheasilvia on top of Veneneia. For this, we use a  $d = 66$  km projectile impacting head-on with a velocity of  $5.4 \text{ km s}^{-1}$ , offset  $40^\circ$  from the centre of Veneneia, and we include pre-impact rotation with period  $P = P_{\text{Vesta}} = 5.3 \text{ h}$  on an axis that goes through the centre of the Veneneia basin. Photo credit for **a**, NASA/JPL-Caltech/UCAL/MPS/DLR/IDA; three-dimensional images were produced using Vapor (www.vapor.ucar.edu).



**Figure 2 | Velocity field lines in snapshots of the simulation of the Rheasilvia impact.** The tangents of the field lines point in the direction of the velocity vectors of the material flow. In the snapshots at times  $t = 800$ ,  $960$  and  $1,200$  s, the velocity field is shown in the co-rotating reference frame, whereas the velocities in the  $t = 7,000$  s snapshot correspond to the inertial frame. The rotation axis and direction are indicated by the black line and arrow, respectively. During crater collapse (shown in the co-rotating images), the large (global) scale of the material flow and the fast spin rate of the body lead to conditions where the Coriolis force becomes significant (low Rossby number, see main text). As a result, the velocity field has a spiral, clockwise pattern, possibly explaining the spiral fracture pattern observed in the Rheasilvia crater on Vesta<sup>9</sup>.

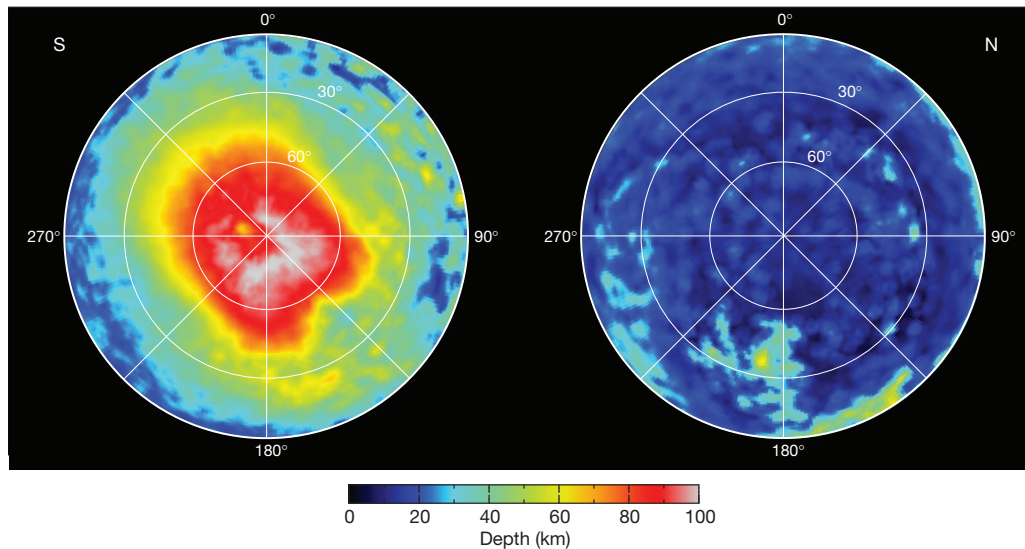
us knowledge of the depth of origin of the surface material. Ejecta are launched from various depths and are globally redistributed on the surface of Vesta; at the same time, material from depth gets exposed within the craters. This leads to maps such as Fig. 3.

Several important facts emerge from the calculations. Most of the ejecta emplaced in the northern hemisphere come from  $\sim 20$  km deep. In the southern hemisphere, rocks exposed in the Rheasilvia area come from  $\sim 60$ – $100$  km, digging beneath the crust, according to interior structure models<sup>21</sup>. In the region where the two basins overlap, material from similar depths is exposed. Note also that after Rheasilvia's formation (that is, about 1 Gyr ago), material has been redistributed by lesser impacts, applying a diffusion effect to what we calculate.

For a given initial structure, the ejecta maps of Fig. 3 can be translated into petrological/mineralogical maps that can be compared to Dawn observations. This provides a direct way to probe the internal petrological structure of Vesta, as well as a means to connect the disparate results concerning Vesta's impact and post-impact evolution. We first tested crustal models<sup>21</sup> with a layer of diogenites covered by an eucritic crust (Fig. 4). For a layered interior, the northern hemisphere should be essentially covered by eucritic ejecta, and the southern hemisphere should display—depending on the initial thicknesses of the layers—variable but large areas of olivine-rich rocks and pure diogenites. These results contrast with the mineralogical and chemical data inferred from Dawn<sup>12,13,26</sup>: although the southern hemisphere is more pyroxene-rich (diogenitic) than the north, most of Vesta's surface appears to be covered by howardites (mixtures of eucrites and diogenites). Pure diogenites have only been reported at two locations in the southern hemisphere<sup>13</sup>.

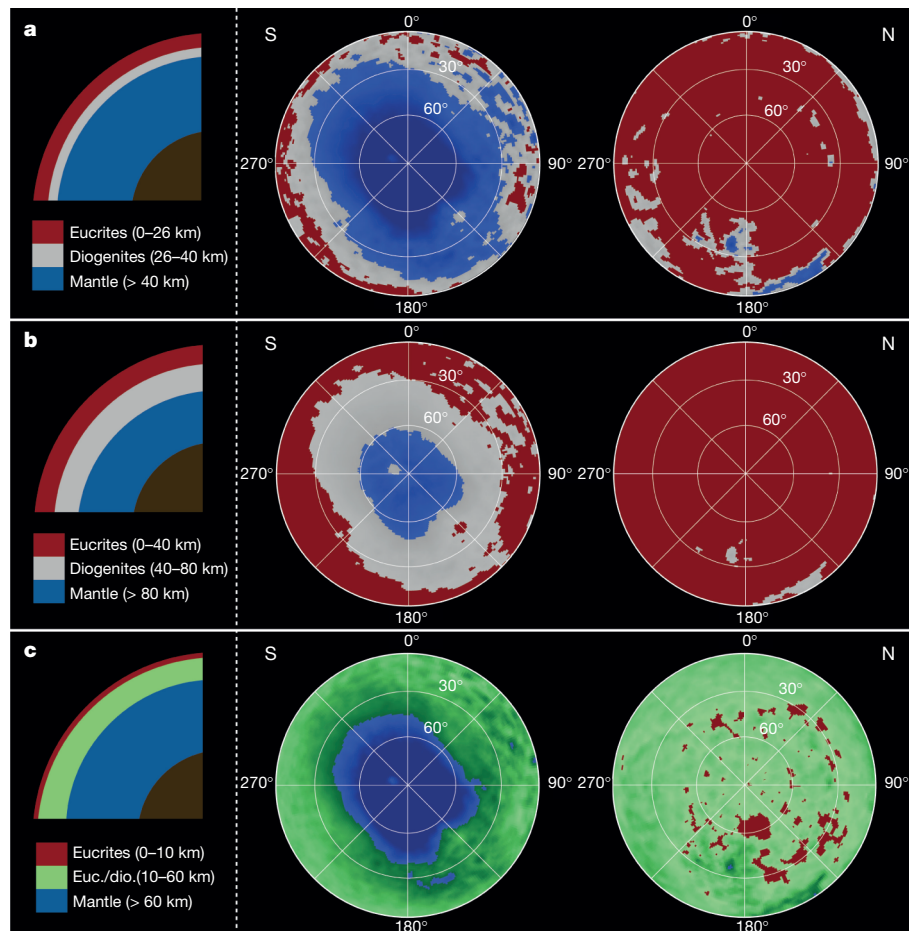
No olivine-rich lithologies have been detected by Dawn<sup>11</sup>. If Vesta is a simple layered magmatic body with structures inherited from the cooling and crystallization of a magma ocean, then these observations are at odds with the large amounts of deep-sourced ejecta generated by our simulations. Because our simulations predict the final topography well, it is likely that material transport is also computed with good accuracy. One possibility is that olivine-rich rocks are actually present on the surface of Vesta but are difficult to detect by current spectroscopic





**Figure 3 | Initial provenance (km depth) of the ejecta and the exposed material on the surface.** Left, southern hemisphere; right, northern hemisphere. The orientation is the same as in the map in Fig. 1. Shown is the final result of the two large impacts forming the Veneneia and Rheasilvia basins at Vesta's south pole. The outcome of the subsequent SPH simulations is shown in a Lambert azimuthal projection (equal area). Assuming a radially stratified

protoplanet, the initial provenance corresponds to the initial depth of the material before the two impacts. The simulations were stopped at  $t = 2 \times 10^4$  s after the impact of the projectile. This time is about equal to Vesta's rotation period ( $P_{\text{Vesta}} = 5.3$  h) and corresponds to  $\sim 10$  dynamical times of  $(G\rho)^{-1/2} \approx 2,000$  s; here  $G$  is the gravitational constant and  $\rho$  is the density.



**Figure 4 | Vesta interior models and the corresponding petrological/mineralogical maps.** **a**, Model 1; **b**, model 2; and **c**, model 3. Models 1 and 2 assume globally distinct layers of eucrites and diogenites<sup>21</sup> (see diagrams at left). Model 3 assumes a more complex structure<sup>24,25</sup>. The colour coding of the material in the diagram at left is used in the maps of ejecta distribution shown centre (southern hemisphere) and right (northern hemisphere). As an initial

target, a solid body of diameter  $d = 550$  km with a mantle (density  $\rho = 2.7 \text{ g cm}^{-3}$ ) and a 240-km-diameter iron core ( $\rho = 7.8 \text{ g cm}^{-3}$ ) was used. The crust is not explicitly modelled, but we track the dynamical evolution and redistribution of crustal layers during the simulation. For each interior model, the compositional maps were then computed using the distribution of ejecta as a function of initial depth (see Fig. 3).

techniques<sup>11,27</sup>. Another alternative is that Vesta has a massive eucritic crust, possibly thicker than 100 km, that is intruded by diogenitic plutons<sup>24,25</sup> (model 3 in Fig. 4). Instead of sampling olivine-rich lithologies at depth, a more diverse source region (eucrite interspersed with diogenite) would be encountered, perhaps better explaining the global howarditic ejecta. Another possibility is that the outer ~100 km of the asteroid was already well mixed by collisions before the Veneneia and Rheasilvia events.

Received 24 September 2012; accepted 4 January 2013.

- McCord, T. B., Adams, J. B. & Johnson, T. V. Asteroid Vesta: spectral reflectivity and compositional implications. *Science* **168**, 1445–1447 (1970).
- McCord, T. B. *et al.* Dark material on Vesta from the infall of carbonaceous volatile-rich material. *Nature* **491**, 83–86 (2012).
- Keil, K. in *Asteroids III* (eds Bottke, W. F. Jr *et al.*) 573–584 (Univ. Arizona Press, 2002).
- Russell, C. T. *et al.* Dawn at Vesta: testing the protoplanetary paradigm. *Science* **336**, 684–686 (2012).
- Thomas, P. C. *et al.* Impact excavation on asteroid 4 Vesta: Hubble Space Telescope results. *Science* **277**, 1492–1495 (1997).
- Asphaug, E. Impact origin of the Vesta family. *Meteorit. Planet. Sci.* **32**, 965–980 (1997).
- Jutzi, M. & Asphaug, E. Mega-ejecta on asteroid Vesta. *Geophys. Res. Lett.* **38**, L01102, <http://dx.doi.org/10.1029/2010GL045517> (2011).
- Ivanov, B. A., Melosh, H. J. & Pierazzo, E. The south pole impact crater on Vesta: numerical modelling. *Lunar Planet. Sci. Conf.* **42**, 1717 (2011); available at <http://www.lpi.usra.edu/meetings/lpsc2011/pdf/1717.pdf>.
- Schenk, P. *et al.* The geologically recent giant impact basins at Vesta's south pole. *Science* **336**, 694–697 (2012).
- Marchi, S. *et al.* The violent collisional history of asteroid 4 Vesta. *Science* **336**, 690–694 (2012).
- McSween, H. Y. *et al.* Dawn's exploration of Vesta's south pole basin — where is the mantle? 75th Annual Meteoritical Society Meeting (2012); available at <http://www.lpi.usra.edu/meetings/metsoc2012/pdf/5020.pdf>.
- De Sanctis, M. C. *et al.* Spectroscopic characterization of mineralogy and its diversity across Vesta. *Science* **336**, 697–700 (2012).
- Reddy, V. *et al.* Color and albedo heterogeneity of Vesta from Dawn. *Science* **336**, 700–704 (2012).
- Jaumann, R. *et al.* Vesta's shape and morphology. *Science* **336**, 687–690 (2012).
- Housen, K. R., Schmidt, R. M. & Holsapple, K. A. Crater ejecta scaling laws: fundamental forms based on dimensional analysis. *J. Geophys. Res.* **88**, 2485–2499 (1983).
- Benz, W. & Asphaug, E. Simulations of brittle solids using smooth particle hydrodynamics. *Comput. Phys. Commun.* **87**, 253–265 (1995).
- Jutzi, M., Benz, W. & Michel, P. Numerical simulations of impacts involving porous bodies. I. Implementing sub-resolution porosity in a 3D SPH hydrocode. *Icarus* **198**, 242–255 (2008).
- Ivanov, B. A. & Melosh, H. J. The Rheasilvia crater on Vesta: numerical modelling. *Lunar Planet. Sci. Conf.* **43**, 2148 (2012); available at <http://www.lpi.usra.edu/meetings/lpsc2012/pdf/2148.pdf>.
- Preusker, F. *et al.* Topography of Vesta from Dawn FC stereo images. *Lunar Planet. Sci. Conf.* **43**, 2012 (2012); available at <http://www.lpi.usra.edu/meetings/lpsc2012/pdf/2012.pdf>.
- Takeda, H. A layered-crust model of a howardite parent body. *Icarus* **40**, 455–470 (1979).
- Ruzicka, A., Snyder, G. A. & Taylor, L. Vesta as the HED parent body: implications for the size of a core and for large-scale differentiation. *Meteorit. Planet. Sci.* **32**, 825–840 (1997).
- Righter, K. & Drake, M. J. A magma ocean on Vesta: core formation and petrogenesis of eucrites and diogenites. *Meteorit. Planet. Sci.* **32**, 929–944 (1997).
- Mittlefehldt, D. W. Petrology and geochemistry of the Elephant Moraine A79002 diogenite: a genotect breccia containing a magnesian harzburgite component. *Meteorit. Planet. Sci.* **35**, 901–912 (2000).
- Barrat, J.-A., Yamaguchi, A., Zanda, B., Bollinger, C. & Bohn, M. Relative chronology of crust formation on asteroid Vesta: insights from the geochemistry of diogenites. *Geochim. Cosmochim. Acta* **74**, 6218–6231 (2010).
- Yamaguchi, A., Barrat, J.-A., Ito, M. & Bohn, M. Post-eucritic magmatism on Vesta: Evidence from the petrology and thermal history of diogenites. *J. Geophys. Res.* **116**, E08009, <http://dx.doi.org/10.1029/2010JE003753> (2011).
- Prettyman, T. H. *et al.* Elemental mapping by Dawn reveals exogenic H in Vesta's regolith. *Science* **338**, 242–246 (2012) (published online, 20 September 2012).
- Beck, P. *et al.* NIR spectral trends of HED meteorites: can we discriminate between the magmatic evolution, mechanical mixing and observation geochemistry effects? *Icarus* **216**, 560–571 (2011).
- Collins, G. S., Melosh, H. J. & Ivanov, B. A. Modeling damage and deformation in impact simulations. *Meteorit. Planet. Sci.* **39**, 217–231 (2004).
- Melosh, H. J. & Ivanov, B. A. Impact crater collapse. *Annu. Rev. Earth Planet. Sci.* **27**, 385–415 (1999).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** M.J. acknowledges support from the Ambizione programme of the Swiss National Science Foundation. E.A. was supported by NASA's Planetary Geology and Geophysics Program. J.-A.B. acknowledges support from the Programme National de Planétologie de l'INSU. W.B. acknowledges support from the Swiss National Science Foundation.

**Author Contributions** M.J. performed and analysed the numerical simulations and led the research. E.A. and W.B. helped to design the numerical study and its scientific formulation. P.G. and J.-A. B. provided the Vesta interior models. P.G. and W.B. initiated the collaboration between the four institutions. All authors contributed to interpretation of the results and preparation of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J. ([martin.jutzi@space.unibe.ch](mailto:martin.jutzi@space.unibe.ch)).



# Hybrid circuit cavity quantum electrodynamics with a micromechanical resonator

J.-M. Pirkkalainen<sup>1</sup>, S. U. Cho<sup>1†</sup>, Jian Li<sup>1</sup>, G. S. Paraoanu<sup>1</sup>, P. J. Hakonen<sup>1</sup> & M. A. Sillanpää<sup>1†</sup>

Hybrid quantum systems with inherently distinct degrees of freedom have a key role in many physical phenomena. Well-known examples include cavity quantum electrodynamics<sup>1</sup>, trapped ions<sup>2</sup>, and electrons and phonons in the solid state. In those systems, strong coupling makes the constituents lose their individual character and form dressed states, which represent a collective form of dynamics. As well as having fundamental importance, hybrid systems also have practical applications, notably in the emerging field of quantum information control. A promising approach is to combine long-lived atomic states<sup>2,3</sup> with the accessible electrical degrees of freedom in superconducting cavities and quantum bits<sup>4,5</sup> (qubits). Here we integrate circuit cavity quantum electrodynamics<sup>6,7</sup> with phonons. Apart from coupling to a microwave cavity, our superconducting transmon qubit<sup>8</sup>, consisting of tunnel junctions and a capacitor, interacts with a phonon mode in a micromechanical resonator, and thus acts like an atom coupled to two different cavities. We measure the phonon Stark shift, as well as the splitting of the qubit spectral line into motional sidebands, which feature transitions between the dressed electromechanical states. In the time domain, we observe coherent conversion of qubit excitation to phonons as sideband Rabi oscillations. This is a model system with potential for a quantum interface, which may allow for storage of quantum information in long-lived phonon states, coupling to optical photons or for investigations of strongly coupled quantum systems near the classical limit.

Superconducting qubits based on Josephson junctions<sup>5</sup> have offered an unparalleled testing ground for quantum mechanics in relatively large systems. At the same time, Josephson devices constitute a promising implementation for quantum information processing. Basic quantum algorithms have been recently demonstrated with phase<sup>9</sup> and transmon<sup>10–12</sup> qubits. The latter operate in the framework of circuit cavity quantum electrodynamics (QED), in which the qubits couple to an on-chip<sup>6</sup> or three-dimensional microwave cavity resonator<sup>13</sup>. The circuit cavity QED set-up, which enables coupling of qubits and non-destructive measurements of quantum states, can be regarded as the most feasible platform for quantum information.

The forthcoming challenges in circuit cavity QED include the construction of an interface for the storage and retrieval of qubit states in a long-lived quantum memory, as well as quantum communication<sup>14</sup> between spatially separated superconducting qubits. Hybrid quantum systems show promise for these purposes because in principle the specific assets of each ingredient can be combined. The merger of macroscopic qubits with spin ensembles is intriguing owing to the long lifetime of the latter<sup>15,16</sup>, but the scheme has the disadvantage of a small coupling at the level of a single atomic degree of freedom.

The quantum regime of micromechanical resonators was first reached only very recently<sup>17,18</sup>. They have been suggested as a plausible interfacial medium for Josephson junction qubits<sup>19–21</sup>, with the potential to achieve the aforementioned results. The first experiment of this kind demonstrated the interaction between a charge qubit and a beam resonating at the frequency  $\omega_m/2\pi \approx 60$  MHz (ref. 22). This experiment,

using a read-out via the micromechanical device, did not allow for time-resolved measurements, which is a prerequisite for coherent state transfer. The resonant coupling was recently demonstrated using a microwave-regime piezoelectric device<sup>23</sup>, with  $\omega_m/2\pi \approx 6.2$  GHz, but the energy decay times were short owing to the high mechanical frequency and special materials required.

Our approach to improving on these issues is to use a membrane-type micromechanical resonator embedded in a complete circuit cavity QED device. This allows access to all the techniques of circuit cavity QED. Notwithstanding that our mechanical resonator can be understood classically, our aim is also to further the understanding of the emergence of ultrastrong coupling of a micromechanical phonon degree of freedom, representing matter, to microwave-frequency light.

On the conceptual level, our device is analogous to an optical cavity QED system where a two-level atom is coupled to two cavities with different frequencies (Fig. 1a). We use a superconducting transmon qubit<sup>8</sup> with capacitance  $C_q$ , single-electron charging energy  $E_C \approx e^2/2C_q$  and Josephson energy  $E_J$  as an artificial two-level system. It is coupled to a highly detuned phononic cavity, formed by a suspended aluminium membrane (Fig. 1d) with flexural-mode resonant frequency  $\omega_m/2\pi \approx 72$  MHz, by means of a position-dependent capacitance  $C_g(x)$ , as well as to a nearly resonant photonic cavity realized as an on-chip coplanar waveguide microwave resonator with frequency  $\omega_c$ .

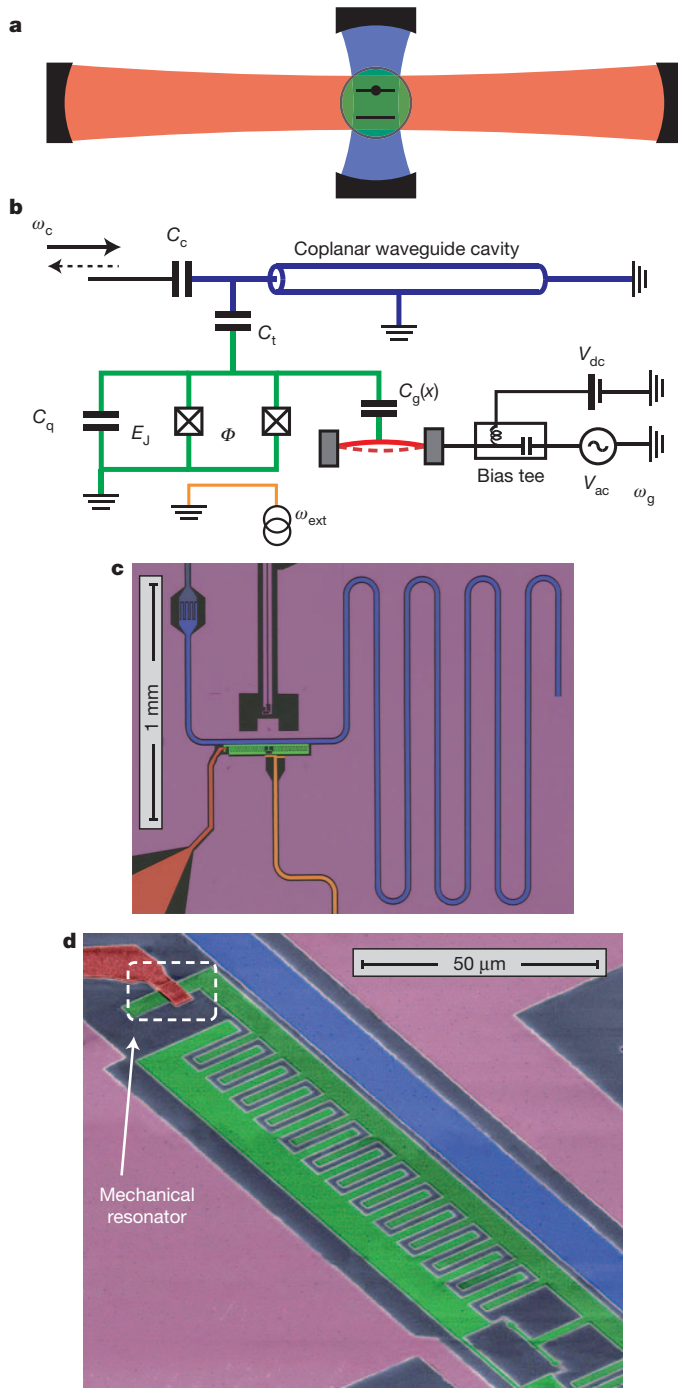
The equivalent electrical circuit (Fig. 1b) for the two-resonator circuit cavity QED system allows the Hamiltonian to be written

$$\hat{H} = \hat{H}_q + \hbar\omega_c(\hat{a}^\dagger\hat{a} + 1/2) + \hbar\omega_m(\hat{b}^\dagger\hat{b} + 1/2) + \hbar g_c(n_0 - \hat{n})(\hat{a}^\dagger + \hat{a}) + \hbar g_m(n_0 - \hat{n})(\hat{b}^\dagger + \hat{b}) \quad (1)$$

where  $\hat{H}_q$  is the qubit Hamiltonian (Supplementary Information, equation (17)),  $\hat{a}$  and  $\hat{b}$  are the annihilation operators of the microwave cavity and of the mechanical resonator, respectively,  $\hat{a}^\dagger$  and  $\hat{b}^\dagger$  are the corresponding creation operators,  $\hat{n}$  is the number operator of Cooper pairs on the transmon island and  $\hbar$  is Planck's constant divided by  $2\pi$ . The interaction between the qubit and the mechanical resonator is given by the electromechanical coupling energy,  $\hbar g_m = x_{zp} V_{dc} (dC_g/dx)(2e/C_q)$ . Here  $x_{zp} \approx 4$  fm is defined as the mechanical zero-point motion,  $V_{dc}$  is a constant voltage applied to the mechanical resonator and  $n_0 = C_g V_{dc}/2e$  is the dimensionless gate charge ( $e$ , electron charge). The coupling between the qubit and the cavity via capacitance  $C_t$ , characterized by the energy  $\hbar g_c$ , allows for a dispersive measurement of the state of the qubit using the state-dependent shift of the cavity frequency<sup>24</sup> (cavity pull).

We use the three lowest-energy eigenstates of the transmon qubit, labelled as the ground state,  $|g\rangle$ , and the first and second excited states, respectively  $|e\rangle$  and  $|f\rangle$ . The corresponding transition frequencies, tunable in the gigahertz regime by a flux bias  $\Phi_{dc}$ , are  $\omega_{g-e}$  and  $\omega_{e-f}$ . We treat  $|e\rangle$  and  $|f\rangle$  as the two-level system. This is somewhat different from the standard concept of a qubit, which usually involves the

<sup>1</sup>Low Temperature Laboratory, Aalto University, PO Box 15100, FI-00076 Aalto, Finland. <sup>†</sup>Present addresses: Korea Research Institute of Standards and Science, Daejeon 305-340, South Korea (S.U.C.); Department of Applied Physics, Aalto University School of Science, PO Box 11100, FI-00076 Aalto, Finland (M.A.S.).



**Figure 1 | Hybrid cavity QED set-up.** **a**, Illustration of a cavity QED system consisting of an atom (green) coupled to two cavities. The (photonic) microwave resonator is coloured blue and the (phononic) lower-frequency mechanical resonator is coloured red. **b**, Circuit schematics of an analogous electromechanical tripartite system, measured using a microwave cavity with a probe tone at frequency  $\omega_c$ . **c**, Optical image of the chip. Shown is the quarter-wave microwave cavity, the superconducting transmon qubit playing the part of the atom, the mechanical resonator with its external control, a flux control line (orange) and the ground plane (magenta). **d**, Scanning electron micrograph showing the 5- $\mu\text{m}$ -long and 4- $\mu\text{m}$ -wide bridge-type mechanical resonator (dashed box) suspended above the qubit island.

ground state, but is beneficial because the phonon coupling grows for higher levels of the qubit. Also, for the same reason, we use the transmon qubit in the regime of relatively small  $E_J/E_C$  ratio.

The coupling of the qubit to the phonons in the mechanical resonator in equation (1) is described by a generic light–matter interaction.

The quantum eigenstates of this coupled system are dressed states, which are combinations of the qubit states and the mechanical resonator Fock states  $|N_m\rangle$ , where  $N_m$  is the phonon number. For large  $N_m$ , the effective coupling,  $g_m\sqrt{N_m+1}$ , becomes comparable to the qubit frequency and the eigenstates attain a highly dressed character, consisting typically of tens of Fock states and of the qubit lowest states (Supplementary Fig. 8).

The qubit transition frequencies are also expected to depend on the number of quanta (a.c. Stark shift). As a phonon-induced effect, it was measured indirectly through the mechanical resonator<sup>22</sup>. The phonon (or mechanical) Stark shift of the frequency  $\omega_{e-f}$  of the  $|e\rangle$ – $|f\rangle$  transition can be calculated beyond the linear regime:

$$\frac{\Delta\omega_{e-f}}{2\pi} = -\frac{\varepsilon_{e-f}}{2} \cos(2\pi n_0) [J_0(2\pi n_x) - 1] \quad (2)$$

Here the argument of the Bessel function  $J_0$  is the motional gate charge amplitude,  $n_x = (hg_m/4E_C)\sqrt{N_m}$ , and the charge dispersion of the transmon qubit is

$$\varepsilon_{e-f} \approx E_C \frac{2^{4m+5}}{m!} \sqrt{\frac{2}{\pi}} \left(\frac{E_J}{2E_C}\right)^{m/2+3/4} \exp(-\sqrt{8E_J/E_C})$$

with  $m = 3$  for this transition. The shift grows first linearly with the phonon number, according to:

$$\Delta\omega_{e-f} \approx \frac{\varepsilon_{e-f}}{2} \pi^2 \cos(2\pi n_0) \left(\frac{hg_m}{4E_C}\right)^2 N_m$$

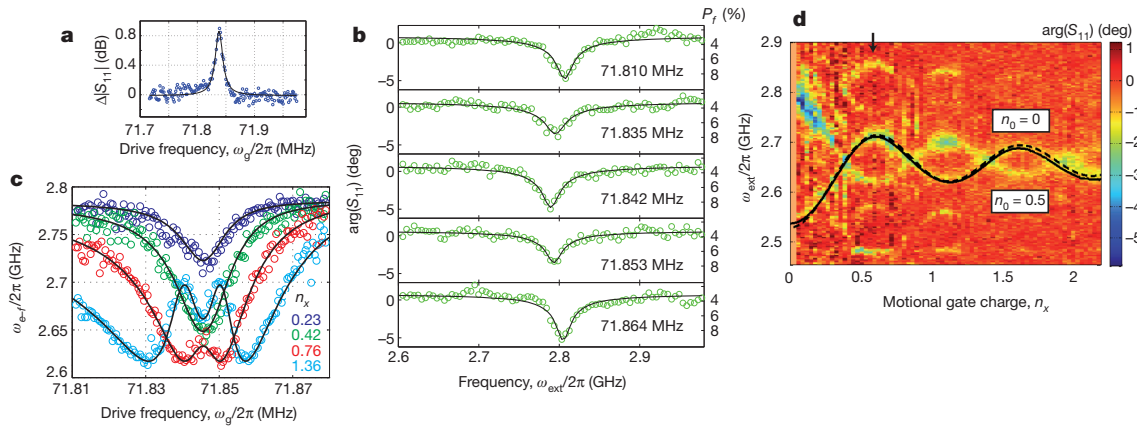
The device (Fig. 1c, d) was fabricated out of aluminium on a sapphire substrate in a process comprising three layers of electron-beam lithography. To suspend the mechanical resonator, we used organic resist as a sacrificial layer<sup>25</sup> (Supplementary Information, section I.A). The quarter-wave microwave cavity has a resonant frequency of  $\omega_c/2\pi = 4.84$  GHz, and capacitively couples to the qubit at the rate  $g_c/2\pi = 100$  MHz. The cavity is coupled to the measurement circuitry by a large input–output capacitor, resulting in the external linewidth  $\gamma_E/2\pi \approx 15$  MHz. The device is mounted in a dilution refrigerator at a temperature of  $T = 25$  mK. Although we did not access the thermal phonon number, we expect the mode to thermalize, as observed for membranes<sup>17</sup>, down to the thermal occupancy  $N_m^T = k_B T/\hbar\omega_m \lesssim 10$ .

By driving the mechanical resonator to a high vibration amplitude, we first identified the lowest mechanical eigenmode of the membrane. The electromechanical interaction was switched on by applying a d.c. voltage  $V_{dc} = 5$  V, which produces an electromechanical coupling  $g_m/2\pi = 4.5$  MHz. The mechanical resonator was driven strongly by an additional a.c. tone combined into a bias-T inside the cryostat (Fig. 1b). When the qubit is biased near the microwave cavity resonance, the cavity absorption in a single-tone measurement is altered when the mechanical mode becomes highly excited. This is due to the qubit being excited to higher levels, which causes changes in cavity pull and absorption. As a result of this, we obtain the mechanical peak at  $\omega_m/2\pi = 71.842$  MHz (Fig. 2a). In what follows, we focus on smaller motion amplitudes.

By virtue of the different frequency ranges of the qubit and the mechanical resonator, we cannot bring them into resonance by tuning the qubit frequency with a flux bias. A highly detuned regime nonetheless allows for complete quantum control, in analogy to what can be done using trapped ions<sup>2,26,27</sup>. The pertinent phenomena are sideband transitions and the Stark shift of the qubit transition frequency.

Two-tone spectroscopy is used in investigating the qubit  $|e\rangle$ – $|f\rangle$  transition to detect the Stark shift. A standard method of probing this transition is to start by applying resonant microwaves to excite the qubit from the ground state,  $|g\rangle$ , to  $|e\rangle$ . However, in our system a small population,  $P_e \approx 25\%$  exists in  $|e\rangle$  even in equilibrium, and so an initial excitation is not necessary. We similarly find  $P_f \approx 5\%$ . The actual excitation microwave tone (frequency,  $\omega_{ext}$ ) is applied to the transmon





**Figure 2 | Mechanical Stark shift.** **a**, The driven mechanical resonance appears as enhanced absorption in single-tone cavity spectroscopy. **b**, In two-tone spectroscopy (flux bias,  $\Phi_{dc}/\Phi_0 = 0.368$ ;  $n_0 = 0.5$ ) the dip representing the transition  $|e\rangle - |f\rangle$  (between the first and second excited states) of the transmon qubit is observed to redshift when the mechanical drive frequency is swept across resonance. Circles denote the measured phase of the cavity reflection, and the solid lines are Lorentzian fits. The scale on the right is the population (probability) in  $|f\rangle$ . With resonant mechanical drive at  $\omega_g/2\pi = \omega_m/2\pi = 71.842$  MHz, mechanical motion is excited with a root mean squared amplitude of 140 fm, which is about 30 times  $x_{ZPF}$ , and the phonon number is

flux coil, and the increased population of level  $|f\rangle$  is distinguished by the phase,  $\arg(S_{11})$ , of the probe tone. This phase carries information about the level populations,  $P_i$ ; here we are interested in enhanced

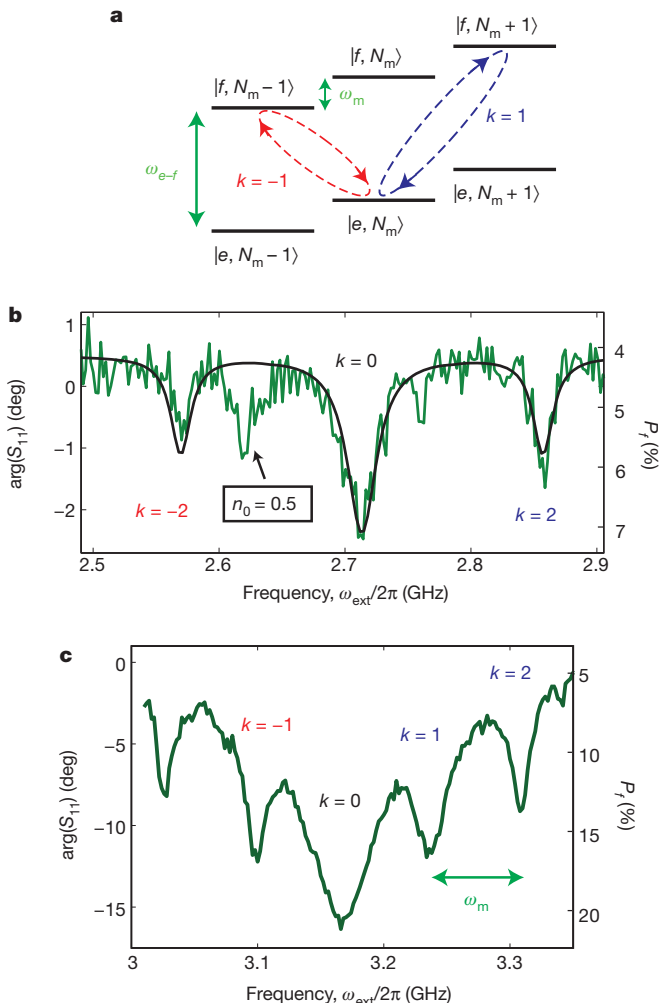
$N_m \approx 10^3$ . In gate charge units, this equals a motional gate charge of  $n_x = 0.09$ . **c**, Shift of the qubit transition at flux bias  $\Phi_{dc}/\Phi_0 = 0.370$  as a function of driving frequency, with four drive amplitudes as labelled. Circles denote the centre of the dip extracted from the data as in **b**, and the solid lines are the modelled response of a mechanical resonator with quality factor  $Q_m = 5,500$ . **d**, Highly nonlinear regime of the mechanical Stark shift with resonant drive at  $n_0 = 0$ . Owing to a single electron fluctuating on and off the island, the transition line corresponding to  $n_0 = 0.5$  is visible as a mirror image. The solid line is a result of full numerical Floquet modelling (Supplementary Information, section II.G), and the dashed line is equation (2).

$P_f$ , characterizing the transition frequency. A shift of  $\Delta\omega_{e-f}/2\pi = -18$  MHz (Fig. 2b) from the bare value,  $\omega_{e-f}/2\pi = 2.81$  GHz, can be attributed to the mechanical Stark shift.

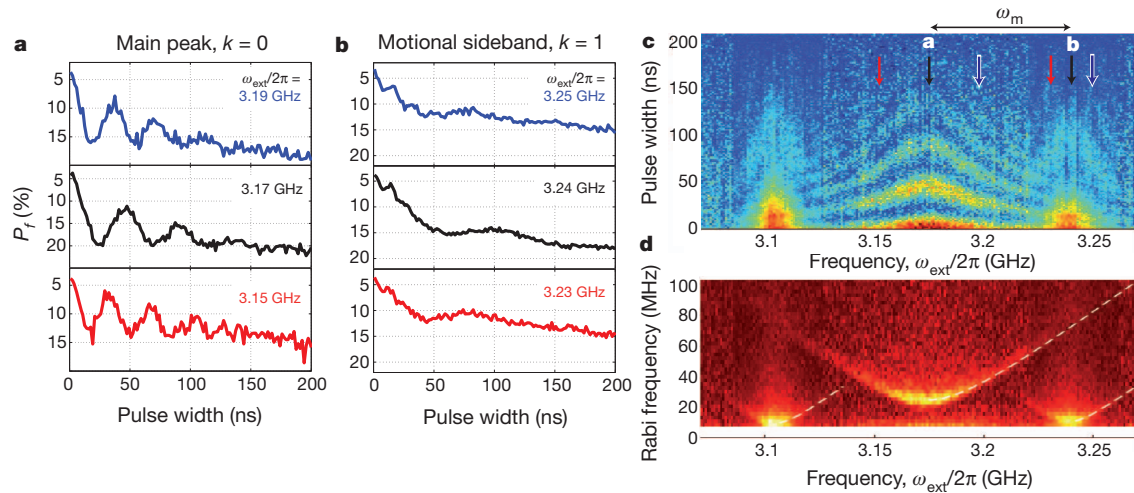
We can extract the centre of the  $|e\rangle - |f\rangle$  transition dip and vary the motion amplitude, as shown in Fig. 2c. The slight difference between the bare value and that in Fig. 2b is due to a flux offset of  $\sim 2m\Phi_0$ . For higher vibration amplitude, the shift becomes more pronounced and eventually becomes highly nonlinear and oscillatory when the motional gate charge reaches values of the order  $n_x \gtrsim 0.5$ . This rather extreme regime can be explored by fixing a resonant mechanical drive, as implemented in the measurement shown in Fig. 2d. This data shows the Bessel-type oscillatory Stark shift, in accord with equation (2).

To model the shift, we first note that the total time-dependent gate charge comprises two components:  $\Delta n_g = n_x + C_g V_{ac}/2e$ . The second term represents cross-talk from the a.c. voltage drive on the mechanical resonator. On resonance, this drive contributes 8% of  $n_x$ . Because the phase of the two terms when the drive is below  $\omega_m$  is opposite to that when the drive is above  $\omega_m$ , they sum up to make the Stark shift asymmetric (Fig. 2c). We used the motional gate charge,  $n_x$ , at the frequency  $\omega_m$  as a common fitting parameter. The analytic result in equation (2) already gives excellent agreement with the measurement (Fig. 2d). We also did a complete numerical calculation using Floquet modelling of the full transmon Hamiltonian (Supplementary Information, section II.G), which well matches the observed spectroscopy (Fig. 2c, d).

The excellent match to theory is further evidence for the mechanical origin of the Stark shift. In addition, the highly nonlinear shift can be



**Figure 3 | Motional sideband transitions.** **a**, In a Stokes (red sideband) scattering event, the qubit exchanges (absorbs or emits) a quantum of frequency  $\omega_{e-f} + k\omega_m$  with the microwave field, whereas the mechanical resonator emits or absorbs  $k$  quanta. In an anti-Stokes (blue sideband) scattering event, the process is reversed. **b**, Cross-cut along the arrow from Fig. 2d at  $n_x \approx 0.6$ . The solid black line is a fit from the analytical model presented in Supplementary Information, equation (49), with decay rate  $\gamma/2\pi = 3$  MHz, dephasing rate  $\gamma_\phi/2\pi = 6$  MHz, and main peak Rabi frequency  $\Omega_0/2\pi = 7$  MHz. **c**, Spectroscopy with the motional gate charge corresponding to the dynamical 'sweet spot' ( $n_x = 0.4$ ; see Supplementary Information, section II.H), where the qubit is insensitive to offset charge fluctuations. The flux bias was  $\Phi_{dc}/\Phi_0 = 0.326$ . The qubit was driven hard at  $\Omega_0/2\pi = 31$  MHz, making the main peak broad.



**Figure 4 | Electromechanical Rabi oscillations.** **a**, Oscillations of the population of the qubit third level,  $|f\rangle$ , measured near the bare qubit frequency,  $k = 0$ . **b**, Time-resolved evolution between phonons and qubit states, near the first blue sideband,  $k = 1$ , with the resonant Rabi frequency  $\Omega_1/2\pi \approx 9.4$  MHz. At the first minimum, around 50–70 ns, the qubit has flipped (at modest fidelity) from  $|e\rangle$  to  $|f\rangle$ , while the mechanical resonator absorbs a quantum. **c**, Qubit population oscillations mapped as a function of the qubit excitation

frequency and Rabi pulse width. An overall decay, visible in **a** and **b** is subtracted for illustration clarity. The traces in **a** and **b** are along the arrows. The decay time constant, related to the total decoherence rate, is  $\sim 70$  ns. **d**, Fourier transform of **c** showing the Rabi frequencies of each sideband growing as a function of excitation detuning. The lines are fits to the expression

$$\Omega_k(\omega_{\text{ext}}) = \sqrt{\Omega_k^2 + (\omega_{e-f} - \omega_{\text{ext}} - k\omega_m)^2}.$$

seen as a manifestation of ultrastrong coupling between light and matter<sup>28,29</sup>. Indeed, the frequency shift due to coupling goes far beyond the linear regime, exceeding the bare mechanical frequency. These effects, however, can be understood by treating the mechanical resonator classically, owing to the high phonon number.

We can use sideband transitions to transfer quanta between the qubit and the mechanical resonator<sup>26,27</sup>. More specifically, the transitions occur between the coupled eigenstates, which are combinations of those of the qubit and of the mechanical resonator. The degree of this dressing depends on the parameter choice. With  $n_x = 1$  in Fig. 2d, the eigenstates involve approximately 20 phonon Fock states and three lowest qubit states with considerable amplitude (0.1–0.4). In Fig. 3c, the effective coupling is small enough that the description in terms of separable qubit and phonon states is valid for the present purpose. Hence, an intuitively appealing picture is to describe them by using uncoupled states  $|e, N_m\rangle$  and  $|f, N'_m\rangle$  (Fig. 3a).

Because the data of Fig. 3b were taken at the gate charge extreme offsets ( $n_0 = 0, 0.5$ ), for symmetry reasons we obtain only two-phonon processes. However, offset drift during data-taking rendered the first-order sidebands visible in Fig. 3b. Because the Rabi frequency is relatively low in Fig. 3b, we can use an analytical result for the qubit population, obtaining excellent agreement with the measured sidebands. However, in Fig. 3c this model, which assumes that the peaks do not overlap, cannot be used. When the motional gate charge is  $n_x \approx 0.4$ , the qubit frequency is insensitive to  $n_0$ , and under this situation we can therefore eliminate both the dephasing due to gate charge fluctuations and the fluctuating single electron (Supplementary Figs 9b and 11).

Although the red and blue sidebands are respectively associated with cooling and heating of the mechanical resonator, at high phonon occupancy, as here, these effects become insignificant. Nevertheless, operations on the motional sidebands are a plausible way of probing the single-phonon regime. Unless the qubit is driven at the sideband resonance, it is effectively decoupled from the mechanical resonator. This idea, complemented by the use of fast flux shifts that tune the qubit frequency, will allow for the coupling to be switched on and off on the nanosecond timescale. We mention that the charge regime used here allows for relatively strong coupling, but possibly makes the qubit susceptible to background charge noise. An approach that avoids this

is to use large  $E_J/E_C$ , for which charge noise vanishes; this requires  $g_m$  to be a few times larger than here (Supplementary Information, section II.I), to reach the quantum regime.

To investigate the energy exchange in the time domain, we performed time-resolved measurements in the set-up of Fig. 3c. Here, instead of using continuous microwave irradiation as in the two-tone spectroscopy of Figs 2 and 3, we pulse the excitation microwave tone at varying widths while monitoring the population of level  $|f\rangle$ . During the pulse, the system is expected to coherently evolve, at the Rabi frequency  $\Omega_k$  (for sideband  $k$ ), between a pair of eigenstates as illustrated by arrows in Fig. 3a. Figure 4 shows the obtained Rabi oscillations on the bare qubit transition ( $k = 0$ ; Fig. 4a) and on the first blue sideband ( $k = 1$ ; Fig. 4b). The latter indicates that the qubit state can swap while simultaneously adding or removing one quantum from the mechanical resonator. The qubit is measured using weak continuous measurement, which here provide the population  $P_f$  immediately following the pulse (Supplementary Information, section I.C). We can map the energy exchange by varying the detuning of the excitation microwave (Fig. 4c), as well as by converting the data to the frequency domain (Fig. 4d). The data shows, as expected, an increase in the Rabi frequency on detuning from a particular sideband. The coherence times are presumably limited by external high-frequency noise, which causes quasiparticle dissipation and a thermal population in the qubit.

The motional sideband operations facilitate the use of the full range of techniques developed in the field of trapped ions<sup>2,26,27</sup> for engineering non-classical states of mechanical resonators or long-distance entanglement of several slowly moving, purely mechanical objects. At the same time, the connection of a superconducting qubit to two different resonators is a prototype of a quantum interface. The microwave resonator allows control over the qubits, whereas the mechanical resonator forms a building block for the conversion of quantum information between microwave light and mechanical motion. The long-lived phonons could be used as a quantum memory. Although here the storage time is limited by the thermal coherence time<sup>17</sup> of the mechanical resonator,  $\tau_T = Q_m/N_m^1\omega_m \approx 1$   $\mu$ s, we note the encouraging recent findings<sup>30</sup> with  $Q_m \approx 10^{10}$  and  $\tau_T$  of the order of seconds. This kind of micromechanical device could further allow for quantum communication by converting phonons into flying photons using radiation pressure coupling, which is naturally suited to a membrane



mirror. An increase in the electromechanical coupling by means of a narrower vacuum gap could bring the interaction between light and true matter all the way into the single-phonon ultrastrong regime.

## METHODS SUMMARY

**Experimental.** The device fabrication includes three layers of electron-beam lithography. The first lithography process patterns everything except the mechanical resonator. Aluminium is deposited by shadow evaporation at thicknesses of 20 and 40 nm, with an oxidation in between to create the Josephson tunnel junctions. The sacrificial layer separating the bridge from the transmon island is defined with poly(methyl methacrylate) used as a negative resist, under high electron dose. The mechanical resonator is defined in the third lithography process. At the end, isotropic O<sub>2</sub> ashing removes the poly(methyl methacrylate) and suspends the bridge. In a tilted scanning electron micrograph, we observe an undulating vacuum gap of about 40–100 nm between the bridge and the qubit island.

With the electromechanical coupling set to zero by  $V_{dc} = 0$ , we first characterized the basic operation of the transmon–cavity circuit QED system. On the basis of single- and two-tone spectroscopy, we determined the following qubit parameters:  $E_{J1}/2\pi = 4.63$  GHz,  $E_{J2}/2\pi = 6.43$  GHz,  $C_q = 61$  fF,  $E_C/2\pi = 318$  MHz. Here,  $E_{J1}$  and  $E_{J2}$  are the Josephson energies of the two junctions, and  $E_J = E_{J1} + E_{J2}$ .

**Theoretical.** In the qubit eigenbasis denoted by the standard Pauli matrices,  $\sigma_x$ ,  $\sigma_y$  and  $\sigma_z$ , the qubit/mechanical resonator Hamiltonian is of the form

$$H \approx -\frac{\omega_e - f}{2} \sigma_z + \omega_m (b^\dagger b + 1/2) + g_{m,z} (b^\dagger + b) \sigma_z + g_{m,x} (b^\dagger + b) \sigma_x$$

The coupling generally contains both diagonal and transverse components, respectively  $g_{m,z}$  and  $g_{m,x}$ . In the charge qubit limit  $g_{m,z} \approx g_m \gg g_{m,x}$ , whereas in the transmon limit  $g_{m,z} \ll g_{m,x} \approx g_m$ . In the present experiment,  $g_{m,z} \approx g_{m,x} \approx g_m$ . In the detuned case, the diagonal coupling with  $g_{m,z}$  is more important. However, in the transmon limit we find that if we make  $g_m/2\pi \gtrsim 25$  MHz by means of a narrower vacuum gap and/or higher d.c. voltage, we can obtain phonon–photon state transfer in the quantum limit within a microsecond.

Received 27 June; accepted 27 November 2012.

- Raimond, J. M., Brune, M. & Haroche, S. Manipulating quantum entanglement with atoms and photons in a cavity. *Rev. Mod. Phys.* **73**, 565–582 (2001).
- Leibfried, D., Blatt, R., Monroe, C. & Wineland, D. Quantum dynamics of single trapped ions. *Rev. Mod. Phys.* **75**, 281–324 (2003).
- André, A. *et al.* A coherent all-electrical interface between polar molecules and mesoscopic superconducting resonators. *Nature Phys.* **2**, 636–642 (2006).
- Nakamura, Y., Pashkin, Y. A. & Tsai, J. S. Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature* **398**, 786–788 (1999).
- Clarke, J. & Wilhelm, F. K. Superconducting quantum bits. *Nature* **453**, 1031–1042 (2008).
- Wallraff, A. *et al.* Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431**, 162–167 (2004).
- Chiorescu, I. *et al.* Coherent dynamics of a flux qubit coupled to a harmonic oscillator. *Nature* **431**, 159–162 (2004).
- Koch, J. *et al.* Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).
- Mariantoni, M. *et al.* Implementing the quantum von Neumann architecture with superconducting circuits. *Science* **334**, 61–65 (2011).
- Reed, M. D. *et al.* Realization of three-qubit quantum error correction with superconducting circuits. *Nature* **482**, 382–385 (2012).

- Fedorov, A., Steffen, L., Baur, M., da Silva, M. P. & Wallraff, A. Implementation of a Toffoli gate with superconducting circuits. *Nature* **481**, 170–172 (2012).
- Dewes, A. *et al.* Quantum speeding-up of computation demonstrated in a superconducting two-qubit processor. *Phys. Rev. B* **85**, 140503 (2012).
- Paik, H. *et al.* Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture. *Phys. Rev. Lett.* **107**, 240501 (2011).
- Ursin, R. *et al.* Entanglement-based quantum communication over 144 km. *Nature Phys.* **3**, 481–486 (2007).
- Zhu, X. *et al.* Coherent coupling of a superconducting flux qubit to an electron spin ensemble in diamond. *Nature* **478**, 221–224 (2011).
- Kubo, Y. *et al.* Hybrid quantum circuit with a superconducting qubit coupled to a spin ensemble. *Phys. Rev. Lett.* **107**, 220501 (2011).
- Teufel, J. D. *et al.* Sideband cooling of micromechanical motion to the quantum ground state. *Nature* **475**, 359–363 (2011).
- Chan, J. *et al.* Laser cooling of a nanomechanical oscillator into its quantum ground state. *Nature* **478**, 89–92 (2011).
- Armour, A. D., Blencowe, M. P. & Schwab, K. C. Entanglement and decoherence of a micromechanical resonator via coupling to a Cooper-pair box. *Phys. Rev. Lett.* **88**, 148301 (2002).
- Tian, L. Entanglement from a nanomechanical resonator weakly coupled to a single Cooper-pair box. *Phys. Rev. B* **72**, 195411 (2005).
- Etaki, S. *et al.* Motion detection of a micromechanical resonator embedded in a d.c. SQUID. *Nature Phys.* **4**, 785–788 (2008).
- LaHaye, M. D., Suh, J., Echterbach, P. M., Schwab, K. C. & Roukes, M. L. Nanomechanical measurements of a superconducting qubit. *Nature* **459**, 960–964 (2009).
- O’Connell, A. D. *et al.* Quantum ground state and single-phonon control of a mechanical resonator. *Nature* **464**, 697–703 (2010).
- Blais, A., Huang, R.-S., Wallraff, A., Girvin, S. M. & Schoelkopf, R. J. Cavity quantum electrodynamics for superconducting electrical circuits: an architecture for quantum computation. *Phys. Rev. A* **69**, 062320 (2004).
- Li, T. F. *et al.* High-frequency metallic nanomechanical resonators. *Appl. Phys. Lett.* **92**, 043112 (2008).
- Monroe, C., Meekhof, D. M., King, B. E. & Wineland, D. J. A Schrödinger cat superposition state of an atom. *Science* **272**, 1131–1136 (1996).
- Roos, C. *et al.* Quantum state engineering on an optical transition and decoherence in a Paul trap. *Phys. Rev. Lett.* **83**, 4713–4716 (1999).
- Niemczyk, T. *et al.* Circuit quantum electrodynamics in the ultrastrong-coupling regime. *Nature Phys.* **6**, 772–776 (2010).
- Tuorila, J. *et al.* Stark effect and generalized Bloch–Siegert shift in a strongly driven two-level system. *Phys. Rev. Lett.* **105**, 257003 (2010).
- Goryachev, M. *et al.* Extremely low-loss acoustic phonons in a quartz bulk acoustic wave resonator at millikelvin temperature. *Appl. Phys. Lett.* **100**, 243504 (2012).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We thank M. Silveri, E. Thuneberg and T. Heikkilä for discussions. This work was supported by the Academy of Finland under CoE in Low Temperature Quantum Phenomena and Devices and project 141559, and by the European Research Council (grant number 240387-NEMSQED) and EU-FP7-NMP-246026. The work benefited from the facilities at the Micronova Nanofabrication Center. J.-M.P. acknowledges support from the Väisälä Foundation, the Emil Aaltonen Foundation and the Kaute Foundation, and J.L. acknowledges support from NGSMP.

**Author Contributions** M.A.S. conceived the experiment and designed the experimental set-up with P.J.H. J.L., J.-M.P. and G.S.P. designed the circuit layout. S.U.C. and J.-M.P. fabricated the samples. J.-M.P. conducted the measurements, developed the theory and wrote the manuscript. S.U.C. helped with the measurements. All authors commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-M.P. ([juha.pirkkalainen@aalto.fi](mailto:juha.pirkkalainen@aalto.fi)).

# Simple organic molecules as catalysts for enantioselective synthesis of amines and alcohols

Daniel L. Silverio<sup>1</sup>, Sebastian Torker<sup>1</sup>, Tatiana Pilyugina<sup>1</sup>, Erika M. Vieira<sup>1</sup>, Marc L. Snapper<sup>1</sup>, Fredrik Haefner<sup>1</sup> & Amir H. Hoveyda<sup>1</sup>

**The discovery of catalysts that can be used to synthesize complex organic compounds by enantioselective transformations is central to advances in the life sciences<sup>1</sup>; for this reason, many chemists aim to discover catalysts that allow for preparation of chiral molecules as predominantly one mirror-image isomer<sup>2</sup>. The ideal catalyst should not contain precious elements<sup>3</sup> and should bring reactions to completion in a few hours through operationally simple procedures. Here we introduce a set of small organic molecules that can catalyse reactions of unsaturated organoboron reagents with imines and carbonyls; the products of the reactions are enantiomerically pure amines and alcohols, which might serve as intermediates in the preparation of biologically active molecules. A distinguishing feature of this catalyst class is the presence of a 'key' proton embedded within their structure. Catalysts are derived from the abundant amino acid valine and are prepared in large quantities in four steps with inexpensive reagents. Reactions are scalable, do not demand stringent conditions, and can be performed with as little as 0.25 mole per cent catalyst in less than six hours at room temperature to generate products in more than 85 per cent yield and  $\geq 97:3$  enantiomeric ratio. The efficiency, selectivity and operational simplicity of the transformations and the range of boron-based reagents are expected to render this advance important for future progress in syntheses of amines and alcohols, which are useful in chemistry, biology and medicine.**

Many biologically active molecules contain nitrogen-substituted carbon stereogenic centres. Routes for efficient preparation of enantiomerically enriched homoallylic amines are therefore of considerable consequence<sup>4</sup>. Anti-cancer agents aza-epothilones A–D<sup>5</sup> (Fig. 1a), leuconicine A and B<sup>6</sup>, natural products that can reverse multi-drug resistance, and immunosuppressant FR235222<sup>7</sup> are among entities the synthesis of which might involve homoallylic amines. Enantioselective addition of an allyl group to an aldimine has thus been the subject of substantial scrutiny<sup>4</sup>. Catalytic protocols have been introduced for preparing homoallylic amines and derivatives with high enantioselectivity; nevertheless, all lack several important attributes. Some demand the intermediacy of allylindiums<sup>8</sup>, prepared *in situ* from allyl halides and the costly metal<sup>9,10</sup>; others entail the use of a rare element<sup>11</sup>. Moreover, the following drawbacks are encountered frequently: difficult-to-access or expensive ligands<sup>12</sup>, high catalyst loadings (for example,  $\geq 10$  mol%)<sup>8–10,12,13</sup>, long reaction times (for example,  $>8$  h)<sup>8–11,13–16</sup>, exceedingly low temperatures (for example,  $-50$  °C or lower)<sup>15,17</sup>, narrow substrate range<sup>9,15,16,18</sup>, and the need for allyltin<sup>11</sup> or moisture-sensitive reagents<sup>13</sup>.

Readily obtainable catalysts for efficient, sustainable and practical enantioselective additions to ketones are equally sought after. Isatins are carbonyl-containing entities that can be converted to enantiomerically enriched 3-hydroxy-2-oxindoles found within alkaloids of substantial biological consequence<sup>19</sup>. Examples are proteasome inhibitors TMC-95A–D with ample potential in the treatment of cancer and immune disorders<sup>20</sup>, and interleukin 6 inhibitor and anti-osteoporosis agent madindoline A (Fig. 1a); proper configuration of the tertiary hydroxyl unit is needed for high activity<sup>21</sup>. A few reports concentrate on catalytic enantioselective allyl additions to isatins; limitations

including the need for toxic tin-based reagents<sup>22</sup>, scarce metal salts<sup>23</sup>, and moderate selectivities<sup>22</sup> complicate these notable advances.

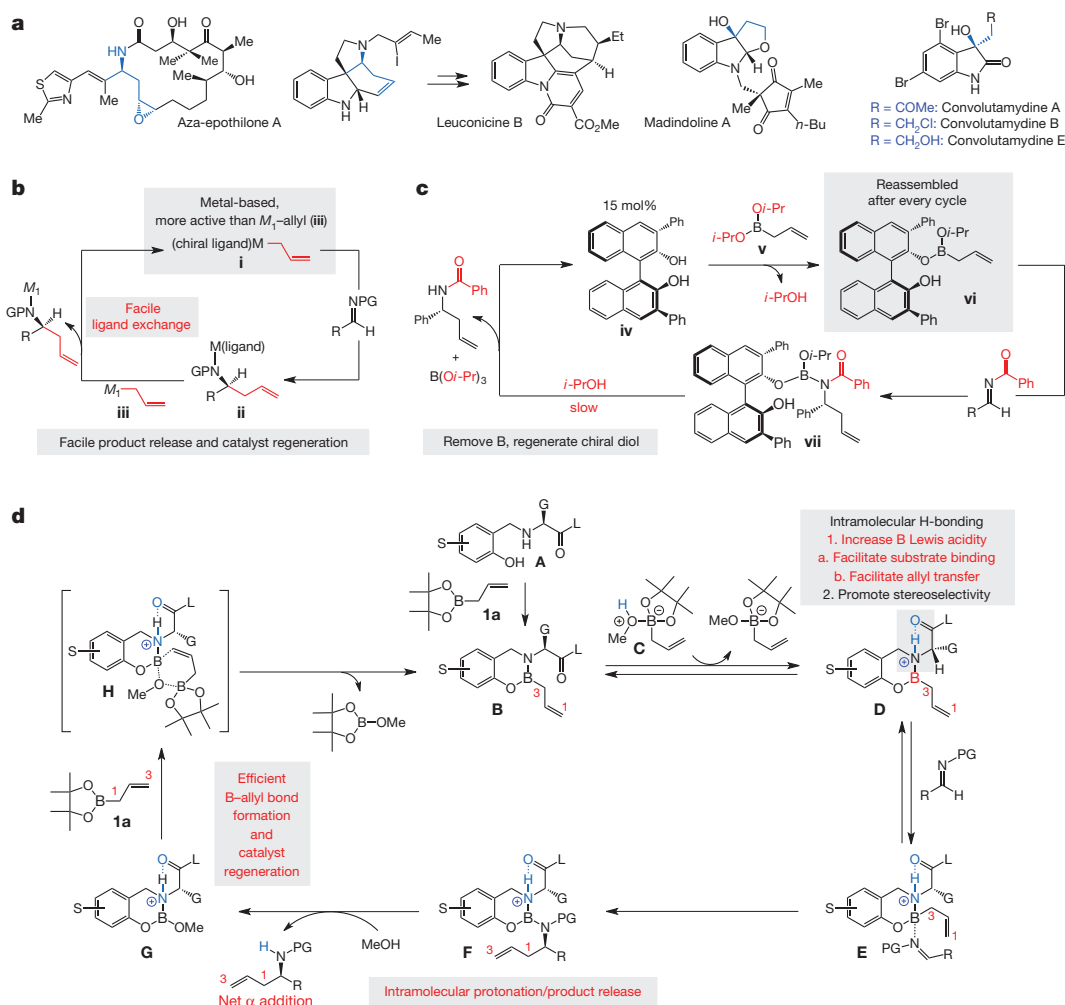
Deliberations regarding catalyst design, alongside consideration of the mechanistic attributes of different extant approaches to catalytic enantioselective allyl additions, led us to opt for metal-free pathways; several factors led to such a conclusion. Most allylmetal reagents are sensitive to oxygen and moisture<sup>24</sup>; their use entails vigilantly controlled conditions. Furthermore, transformations with  $\pi$ -allylmetal complexes are usually either not diastereoselective<sup>9,17</sup> or one possible isomer remains inaccessible<sup>13,23</sup> regardless of whether the *E* or *Z* allylic reagent is employed. Although strategies involving stoichiometric quantities of enantiomerically pure substrates offer stereoselective alternatives, the transformations suffer from similar limitations (see the Supplementary Information for bibliography).

There is one metal-free catalytic method for enantioselective additions of allylboron reagents to imines<sup>13</sup>; reactions, however, proceed less readily and demand higher catalyst amounts and longer reaction times than when allylmetal species are involved (Fig. 1c); additionally, moisture-sensitive allylboron derivatives are required (v, Fig. 1c), and similar to transformations with crotylmetal reagents, only one product diastereomer can be prepared<sup>13</sup>. Reactions with metal-containing systems are probably more efficient because of swift ligand exchange leading to fast catalyst regeneration (Fig. 1b): the swap between a homoallyl metal-amide (ii) and an allyl reagent (iii) to re-form the active complex i can occur rapidly. In contrast, allylboron vi needs to be re-assembled after each cycle (Fig. 1c): the enantiomerically enriched vii must first be converted to diol iv by protolytic removal of the boron and product moieties; the diol then reacts with allylboron v to regenerate vi. Mechanistic studies indicate that it is indeed the regeneration of the diol iv that hampers reaction rate<sup>25</sup>. Thus, to ensure re-formation of vi, a more reactive but moisture sensitive allylboron (v) was prepared and used.

The above analysis led us to conclude that a pathway must be conceived such that the catalyst is reproduced rapidly but without a sensitive allylboron and the benefit of the favourable kinetics available to metal-containing intermediates. Accordingly, we drafted the blueprint outlined in Fig. 1d. A substituted aminophenol (A) offered an attractive possibility; such molecules are structurally modular and synthesized by dependable manipulations; chiral allylboron B would be generated by reaction with the relatively robust (pinacolato)allylboron 1a. At this juncture, several challenges become evident: (1) the boron centre, bearing a comparatively electron-donating amine ligand, would have to be rendered sufficiently Lewis acidic to bind readily with the substrate. (2) The stereogenic centre resides at a conformationally mobile arm of the catalyst; high enantioselectivity would demand strong differentiation between the diastereotopic faces of the coordinated imine. (3) A mechanism for quick catalyst regeneration would have to be identified. We considered that a solution could involve an internal hydrogen bond, bridging the catalyst's amide carbonyl and boron-bound nitrogen (Fig. 1d). Such electrostatic attraction would increase the boron centre Lewis acidity to facilitate substrate binding ( $\rightarrow$ E) and C–C bond formation ( $\rightarrow$ F) and rigidify

<sup>1</sup>Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, Massachusetts 02467, USA.





**Figure 1** | The significance of homoallylic amines and alcohols illustrated by three approaches to their catalytic enantioselective synthesis.

**a**, Representative biologically active natural products synthesized via chiral homoallylic amines and 3-hydroxy-2-oxindoles. **b**, With a metal-containing catalyst, high rates are achieved through facile ligand exchange; catalytic allyl addition is shown. **c**, In a metal-free system, the catalyst must be reassembled

before each cycle. **d**, A boron-based small-molecule catalyst might be designed such that electrostatic forces caused by amine protonation lead to fast reaction rates and high enantioselectivities. Catalytic cycles deliver net  $\alpha$  addition ( $C_1-B \rightarrow C_1-C$ ) resulting from two  $\gamma$ -selective processes ( $G \rightarrow B$  and  $D \rightarrow F$ ). Facile catalyst regeneration may occur through allylation of **G** via **H**. PG, protecting group.

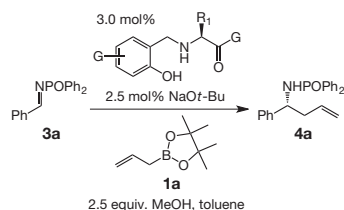
the catalyst•substrate complex **E**, engendering high enantioselectivity. Turnover could be facilitated by acceleration of product release through intramolecular protonation in **F**, affording the desired product and chiral intermediate **G**, which could then react to regenerate the catalytically active **B** through a structure such as **H** (Fig. 1). As will be detailed below, the facility with which **H** can be accessed is central to the high turnover rates achieved; the intermediacy of **H** has stereochemical consequences that are among the hallmarks of the present system. We further noted a significant implication of the projected mechanistic scenario: once the boron-based catalyst (**B**) is generated (that is, after the first cycle), subsequent cycles would deliver net  $\alpha$  addition of an allyl unit ( $C_1-B \rightarrow C_1-C$ ) resulting from two  $\gamma$ -selective processes (that is,  $G \rightarrow B$  and  $D \rightarrow F$ ; Fig. 1d).

We first probed the ability of aminophenols **2a–2h** (Table 1, Fig. 2) to serve as catalyst precursors for reactions involving commercially available (pinacolato)allylboron **1a** and *N*-phosphinoylimine **3a** (Fig. 2). The choice of the *N*-activating group, despite its ostensible non-optimal atom economy, was for several reasons. The derived imines, aryl- or alkyl-containing, can be prepared efficiently; such entities are relatively robust and generate products that are easy to purify owing to their strong tendency to be crystalline (chromatography avoided). There are inexpensive and efficient mildly acidic methods for removal of the phosphorous-based protecting group

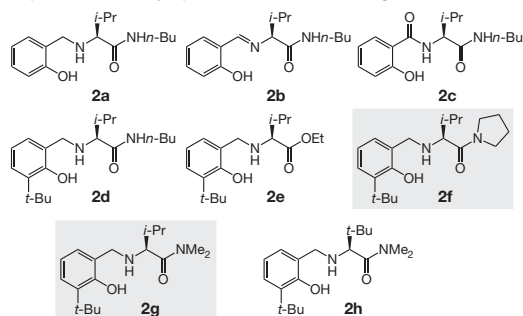
and generation of the parent amines<sup>24,26</sup>. Such protocols tolerate many commonly used functional groups and do not require strongly reductive conditions (for example, diisobutylaluminium hydride<sup>13</sup> required in Fig. 1c), or costly metal salts (for example, SmI<sub>2</sub>; refs 18, 27) and/or alkyl lithium reagents<sup>10</sup>.

When imine **3a** and allylboron **1a** are subjected to 3.0 mol% amino alcohol **2a** (Fig. 2, Table 1, entry 1), 2.5 mol% NaOt-Bu and 2.5 equiv. of MeOH, 71% conversion to homoallylamide **4a** is observed in four hours (with an enantiomeric ratio, e.r., of 74.5:25.5). With Schiff base **2b** or amide **2c** (Table 1, entries 2–3), there is minimal transformation. Placement of a sizeable *t*-butyl unit adjacent to the phenol group in **2d** (Table 1, entry 4), incorporated with the idea of discouraging dimerization of two or more amino alcohols in solution, led to improved efficiency (>98% conversion); the superior enantioselectivity (91:9 e.r.) reflects a substantially more facile process initiated by the chiral catalyst, because control experiments indicate that allyl addition proceeds with reasonable efficiency in its absence (70% conversion, 75 min, 22 °C). Lower e.r. and diminished reactivity is furnished by less Lewis basic ethyl ester **2e** (Table 1, entry 5). With dialkylamides **2f** and **2g** (Table 1, entries 6 and 7) additions proceed to completion readily, affording the desired amide in approximately 96:4 e.r. Reaction with **2h** (Table 1, entry 8) is more selective (98:2 e.r.) but requires the exorbitantly expensive *t*-Leu residue. Lastly, similar

The representative transformation:



Representative catalyst precursors used in screening studies:



**Figure 2** | Chiral amino alcohols as candidates for catalyst precursors. Top, the representative transformation; bottom, the representative catalyst precursors. Table 1 reports the results. The lack of activity shown by catalysts derived from **2b** and **2c** is consistent with the mechanistic scenario outlined in Fig. 1d, as the requisite chiral allylboron species **D** cannot be generated. Also consistent is the low activity and enantioselectivity by ester-containing **2e**, underscoring the pivotal role of the catalyst's Lewis basic C terminus in establishing a hydrogen bond.

efficiency and enantioselectivity is attained when organic amines are used as base (for example, 1,8-diazabicycloundec-7-ene, dbu).

A wide array of imines undergoes allyl additions with 3.0 mol% of amino alcohol **2g** and 1.5 equiv. of allylboron **1a** within six hours at ambient temperature (Table 2). Homoallylamides, including those that bear heterocyclic moieties, such as a furyl or a pyridyl unit (entries 11 and 12, Table 2), are isolated often in >85% yield and ≥97:3 e.r. As the syntheses of **4m** and **4n** illustrate (Fig. 3), use of 2-substituted allylboron reagents results in equally efficient and enantioselective processes. The method can be extended to additions with alkenyl-, alkynyl- and alkyl-substituted aldimines (Table 3).

Stereochemical models, supported by computational studies (Supplementary Information), are presented in Fig. 3. Association of the *N*-phosphinoylimine with the boron centre and allyl addition takes place as depicted (**I**→**II**); the allyl and the *i*-Pr groups of the catalyst in **III** bring about steric repulsion. The proposed scenario assigns an additional role to hydrogen-bonding: a three-pronged association involving the catalyst's amine and amide carbonyl and the phosphinoyl unit is established; reactions with *N*-aryl imines, which lack an appropriate hydrogen-bond acceptor, engender minimal enantioselectivity. Other observations support the hypothesis regarding the function of electrostatic interactions (Fig. 1d). Kinetic studies point to the C–C bond forming step as rate determining, with imines bearing electron-donating groups reacting at a slower pace (Supplementary Information). There is <2% conversion without MeOH. Furthermore, treatment of a solution of **2g** with one equivalent of NaOt-Bu results in rapid and complete phenol deprotonation. The addition of five equivalents of MeOH does not lead to major changes; when two equivalents of allylboron **1a** are introduced, allowing for Lewis acid activation of the alcohol additive (Fig. 1d), the phenol is regenerated immediately (>98%). The above observations support the notion that, overall, the mixture is a buffered acidic solution.

Several key features of the catalytic system are outlined in Fig. 4. The chiral amino alcohol has the low molecular weight of 306.4 g mol<sup>−1</sup>; it can be prepared on a multi-gram scale by an uncomplicated four-step sequence involving valine, inexpensively available as either enantiomer,

**Table 1** | Examination of various amino alcohols

Entry number	Amino alcohol; (mol%)	Time (h); T (°C)	Conversion (%) <sup>*</sup>	Enantiomeric ratio <sup>†</sup>
1	<b>2a</b> ; 3.0	4.0; 22	71	74.5:2.5
2	<b>2b</b> ; 3.0	4.0; 22	<2	ND
3	<b>2c</b> ; 3.0	4.0; 22	<10	ND
4	<b>2d</b> ; 3.0	4.0; 22	>98	91:9
5	<b>2e</b> ; 3.0	4.0; 22	47	80:20
6	<b>2f</b> ; 3.0	4.0; 22	>98	96:4
7	<b>2g</b> ; 3.0	4.0; 22	>98	96.5:3.5
8	<b>2h</b> ; 3.0	4.0; 22	97	98:2

The representative transformation is shown in Fig. 2. Reactions were carried out in toluene under an atmosphere of nitrogen gas. ND, not determined.

<sup>\*</sup> Conversion to the desired product as measured by analysis of 400 MHz <sup>1</sup>H NMR spectra of unpurified mixtures versus an internal standard of 9-methylantracene; the variance of values is estimated to be <±2%.

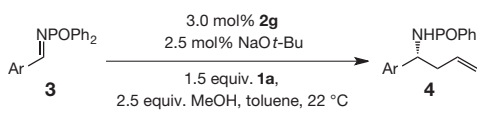
<sup>†</sup> Enantiomeric ratios were determined by HPLC analysis; the variance of values is estimated to be <±2%. See Supplementary Information for details.

and other cheap materials. Purification of **2g**, indefinitely stable to air and moisture, entails routine filtration without the need for costly chromatography procedures. Enantioselective additions are scalable, as the case in Fig. 4a illustrates; reaction work-up is no more than solvent evaporation (analytically pure homoallylamine is obtained by trituration)—distillation or silica gel chromatography is, again, not needed. Such a simple and cost effective product isolation procedure (no need for expensive chromatography solvents) is largely due to the diphenylphosphinoyl unit, more than compensating for its perceived lack of atom economy.

Congruent with the pathway in Fig. 1d, and confirmed by the reaction with d<sub>2</sub>-**1a** (Fig. 4b), the overall transformation takes place with net α selectivity (d<sub>2</sub>-**4o**; 95% α)<sup>27</sup>. Homoallylic amide **4o** can be used in enantioselective synthesis of anti-cancer agents aza-epothilones (Fig. 1a)<sup>5</sup>. The ability to convert the C–B of an allylboron entity to a C–C bond, while generating a *N*-substituted stereogenic centre, has critical implications vis-à-vis its utility in stereoselective synthesis. With allylboron **S-9** or its enantiomer **R-9**, accessed in 94:6 e.r. by a Cu-catalysed protocol<sup>28</sup>, homoallylamides **10** and **11** are obtained in 84% and 93% yield, 84:16 and 83:17 diastereomeric ratio (d.r.), respectively, and 95:5 e.r. (for the major diastereomer); reaction with allylboron **12**<sup>28</sup>, bearing a quaternary carbon stereogenic centre (95:5 e.r.), delivers **13** in 70% yield (pure diastereomer), 89:11 d.r. and 95:5 e.r. (major isomer) (Fig. 4c). Alternative diastereomeric products can be synthesized through the use of the other enantiomer of a chiral allylboron (**10** versus **11**, Fig. 4c). There is complete α selectivity in all instances. The route charted in Fig. 1d implies that a net γ-selective addition should result from the initial catalytic cycle (that is, boron-based catalyst **B** first generated by ligand exchange); that none of the homoallylamine from overall γ-addition is detected suggests that the catalyst is derived from a minute fraction of the amino alcohol, or **B** is initially formed by a pathway to be elucidated. Reaction with sterically demanding **12**, for reasons that remain to be determined, proceeds more readily when performed with Zn(Ot-Bu)<sub>2</sub>.

The reversal in the stereochemical identities in the reactions shown in Fig. 4c, ascertained through X-ray crystallography, supports the suggested general mechanism and the pivotal allyl exchange step leading to rapid catalyst regeneration. As initially put forth in Fig. 1d, stereoselective formation of **13** begins with product release by intramolecular proton transfer (Fig. 4d), leading to the formation of **VI**, wherein the boron centre is stabilized by chelation with the Lewis basic amide group. Subsequent reaction with MeOH yields **VII** (Fig. 4d). Stereoselective generation of chiral allylboron species **IX** can proceed via **VIII** (Fig. 4d), involving a synclinal (cyclic) transition state<sup>4</sup>; otherwise, the corresponding *Z* isomer of **IX** or a mixture of the two would be formed and the reverse diastereoselectivity or little stereochemical preference would be observed. Selective formation of **13** would take place through transition complex **X**.

The catalytic strategy can be applied to carbonyl-containing substrates, entities that do not readily lend themselves to chiral auxiliary

**Table 2 | Catalytic enantioselective allyl additions to aryl-substituted imines**


Entry number	Ar	Time (h)	Conversion (%) <sup>*</sup> ; yield (%) <sup>†</sup>	Enantiomeric ratio <sup>‡</sup>
1	Ph; <b>3a</b>	4.0	>98; 95	96.5:3.5
2	<i>o</i> -FC <sub>6</sub> H <sub>4</sub> ; <b>3b</b>	4.0	>98; 91	98:2
3	<i>o</i> -BrC <sub>6</sub> H <sub>4</sub> ; <b>3c</b>	4.0	>98; 86	97.5:2.5
4	<i>o</i> -MeC <sub>6</sub> H <sub>4</sub> ; <b>3d</b>	6.0	>98; 91	93.5:6.5
5	<i>m</i> -BrC <sub>6</sub> H <sub>4</sub> ; <b>3e</b>	4.0	>98; 95	98:2
6	<i>p</i> -BrC <sub>6</sub> H <sub>4</sub> ; <b>3f</b>	6.0	>98; 91	97:3
7	<i>p</i> -CF <sub>3</sub> C <sub>6</sub> H <sub>4</sub> ; <b>3g</b>	4.0	>98; 93	98:2
8	<i>p</i> -MeO <sub>2</sub> C <sub>6</sub> H <sub>4</sub> ; <b>3h</b>	4.0	>98; 92	98:2
9	<i>p</i> -MeOC <sub>6</sub> H <sub>4</sub> ; <b>3i</b>	4.0	>98; 98	96.5:3.5
10	<i>p</i> -( <i>n</i> -Bu) <sub>2</sub> C <sub>6</sub> H <sub>4</sub> ; <b>3j</b>	4.0	95; 93	92:8
11	2-furyl; <b>3k</b>	6.0	>98; 93	98:2
12	3-pyridyl; <b>3l</b>	4.0	90; 75	98:2

Reactions were carried out in toluene under an atmosphere of nitrogen gas.

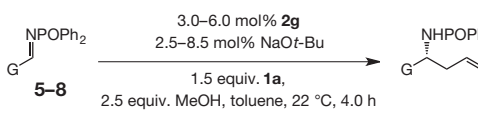
<sup>\*</sup> Conversion to the desired product as measured by analysis of 400 MHz <sup>1</sup>H NMR spectra of unpurified mixtures versus an internal standard of 9-methylanthracene; the variance of values is estimated to be  $\pm 2\%$ .

<sup>†</sup> Yield of isolated product after purification; the variance of values is estimated to be  $\pm 2\%$ .

<sup>‡</sup> Enantiomeric ratios were determined by HPLC analysis; the variance of values is estimated to be  $\pm 2\%$ . See Supplementary Information for details.

approaches. The catalyst derived from **2g** promotes efficient enantioselective reactions with isatins, potential precursors to tertiary alcohols used in drug development<sup>19</sup>. With 0.5–2.0 mol% **2g** and 1.5 equiv. of the allylboron reagent, addition to N-protected isatins is complete at 22 °C within two hours (Fig. 5a); homoallylic alcohols are obtained in 84–98% yield and 91.5:8.5–98.5:1.5 e.r. As the syntheses of **15a** and **15b** exemplify, enantioselective allyl addition/amide deprotection can be carried out in a single vessel easily and with exceptional efficiency. Homoallyl carbinol **15a** is applicable to the synthesis of madindoline A<sup>29</sup> and **15b** is a potential intermediate en route to different convolutamydines (Fig. 3a)<sup>30</sup>. A stereochemical model similar to that offered for additions to imines applies (XI and XII, Fig. 5b). Allyl addition to acetophenone under the same conditions proceeds with high efficiency (3.0 mol% **2g**, >98% conversion in 4.0 h) but in 70:30 e.r., consistent with the proposed mechanistic model.

Another readily accessible organoboron reagent may be used in the present set of catalytic transformations: in the presence of 0.5 mol%

**Table 3 | Catalytic enantioselective allyl additions to alkenyl-, alkynyl- and alkyl-substituted imines**


Entry number	G	2g (mol%); NaOt-Bu (mol%)	Conversion (%) <sup>*</sup> ; Yield (%) <sup>†</sup>	Enantiomeric ratio <sup>‡</sup>
1		3.0; 2.5	>98; 84	>99:1
2		3.0; 2.5	>98; 95	>99:1
3		3.0; 2.5	>98; 98	>99:1
4		3.0; 2.5	>98; 96	98:2
5		2.5; 2.5	>98; 96	98:2
6		3.0; 2.5	>98; 95	88:12
7		6.0; 5.0	66; 50	>99:1
8		6.0; 5.0	70; 51	>99:1
9		6.0; 8.5	90; 71	97.5:2.5

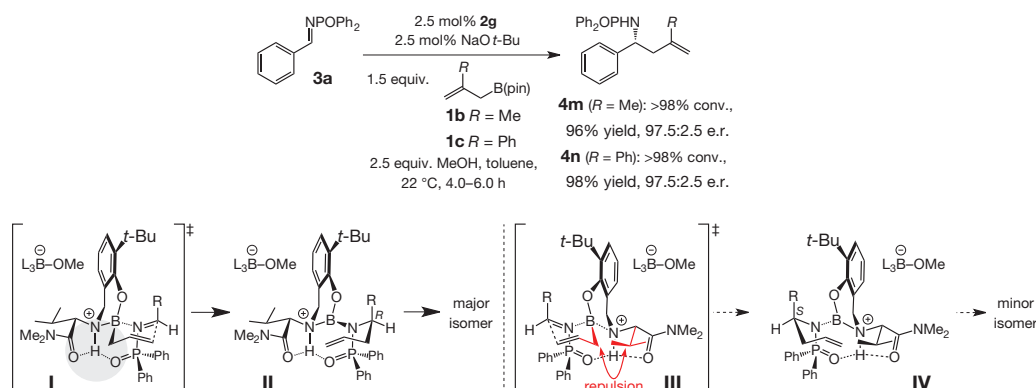
Reactions were carried out in toluene under an atmosphere of nitrogen gas.

<sup>\*</sup> Conversion to the desired product as measured by analysis of 400 MHz <sup>1</sup>H NMR spectra of unpurified mixtures versus an internal standard of 9-methylanthracene; the variance of values is estimated to be  $\pm 2\%$ .

<sup>†</sup> Yield of isolated product after purification; the variance of values is estimated to be  $\pm 2\%$ .

<sup>‡</sup> Enantiomeric ratios were determined by HPLC analysis; the variance of values is estimated to be  $\pm 2\%$ . See Supplementary Information for details.

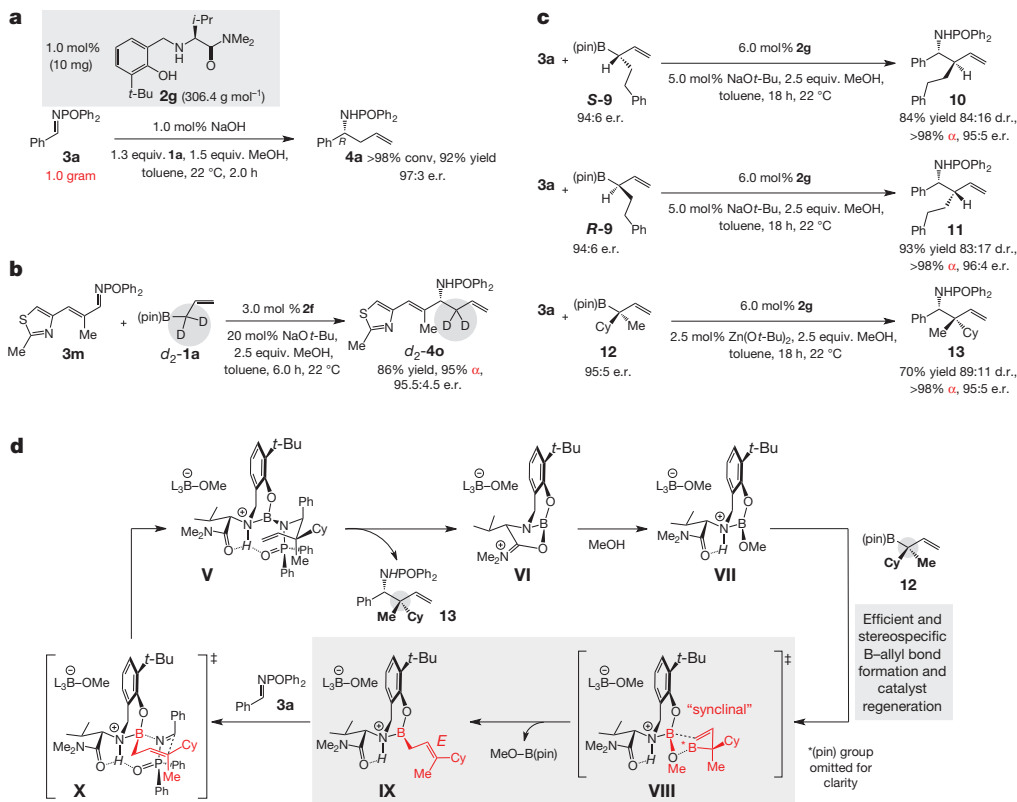
**2g**, reaction of benzyl amide **14c** or *p*-methoxybenzyl amide **14d** with commercially available (pinacolato)allylboron **19** is complete within four hours at ambient temperature, affording allyl carbinols **20a** and **20b** in 98:2 and 96:4 e.r. and 91% and 90% yield, respectively (Fig. 5c). Similar to the reaction with **14d**, addition to silylamide **14a** can be performed on the gram scale in a standard fume hood with 0.25 mol% **2g** and 1.05 equiv. of **19**; C–C bond formation is complete within two minutes and the silyl group is removed through mild acidic workup to afford **21**, which can be isolated in high purity without chromatography, in 90% overall yield and >99:1 e.r. The enantioselective synthesis of  $\alpha$ -hydroxy alcohol **22** further demonstrates utility;



**Figure 3 | Efficient and enantioselective catalytic allyl additions to aldimines.** Aryl-, alkenyl-, alkynyl- and alkylimines can be used to generate homoallylic amides with high efficiency and enantioselectivity (Tables 2 and 3). Top, use of 2-substituted allylboron reagents results in equally efficient and enantioselective processes. Bottom, mechanistic models account for the observed enantioselectivity and involve hydrogen-bonding interactions that bring the reaction components together, promote high enantiotopic face

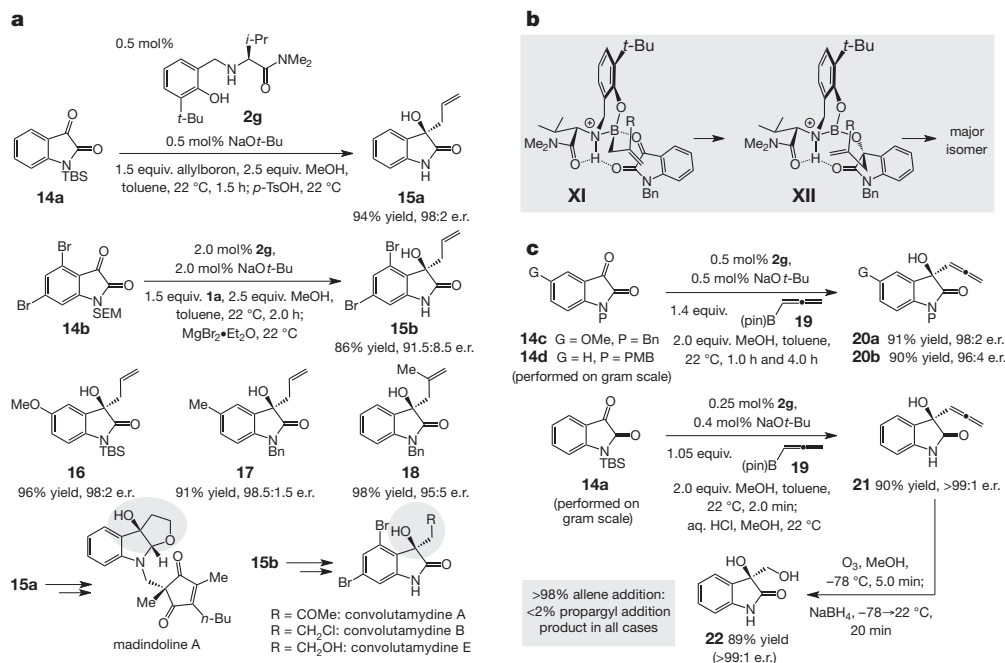
differentiation by enforcing an organized transition structure, and facilitate bond formation by minimizing electron–electron repulsion caused by the converging heteroatoms (see text for discussion). This model is supported by the X-ray crystal structures of **2g** and its HCl salt, which contain a proton-bridge connecting the amine and carbonyl units (see Supplementary Information).





**Figure 4 | Practical, scalable and highly  $\alpha$ -selective catalytic enantioselective allyl additions to imines.** **a**, A practical catalytic protocol. Amino alcohol **2g** is prepared in multi-gram quantities inexpensively by simple procedures; additions are easily performed on gram scale. Key points: **2g** is prepared in four steps in 73% yield (~5 g scale from inexpensive materials, with no chromatography, and is indefinitely air-stable at 22 °C; the reaction is performed in a common fume hood with commercial (undistilled/unpurified) allylboron, NaOH, MeOH, without aqueous extraction or silica gel chromatography. **b**, Deuterium-labelling experiments support the preference for high  $\alpha$  selectivity. Shown is synthesis of a fragment of aza-epothilone A. **c**, Various attributes of the chiral catalyst allow access to homoallylamides with an additional tertiary or quaternary carbon stereogenic centre with high  $\alpha$ -, diastereo- and enantioselectivity. Shown are reactions with  $\alpha$ -substituted allylborons, with

reversal of stereochemistry at the B-substituted carbon. **d**, The stereochemical outcome with substituted allylboron reagents support the proposed mechanism and shed light on the efficient and stereoselective allyl transfer phase of the catalytic cycle (catalyst regeneration/product release). Facile allyl transfer and catalyst regeneration is pivotal to high catalyst efficiency. Hydrogen-bonding in **VII** stimulates enhanced Lewis acidity at the chiral catalyst's boron centre, favouring donation by the  $\pi$  bond of the organoboron reagent **12** (see **VIII**), facilitating stereoselective generation of **IX**. Conversions and diastereomeric ratios were measured by analysis of 400 MHz  $^1\text{H}$  NMR spectra of unpurified mixtures; the variance of values estimated to be  $< \pm 2\%$ . Yields correspond to isolated and purified products ( $\pm 2\%$ ). Enantiomeric ratios were determined by HPLC analysis ( $\pm 2\%$ ). See Supplementary Information for experimental details and spectroscopic analyses.



**Figure 5 | Catalytic enantioselective additions to isatins and reactions with an allenylboron reagent.** **a**, Catalytic enantioselective allyl additions to isatins afford homoallylic alcohols. **b**, A stereochemical model proposed to account for the enantioselectivities. **c**, Broad applicability is illustrated by enantioselective allene additions to isatins, performed with commercially available organoboron reagent **19**. All reactions were carried out in toluene under an atmosphere of nitrogen gas. Conversions measured by analysis of 400 MHz <sup>1</sup>H NMR spectra of unpurified mixtures; the variance of values estimated to be  $\leq \pm 2\%$ . Yields correspond to isolated and purified products ( $\pm 2\%$ ). Enantiomeric ratios were determined by HPLC analysis ( $\pm 2\%$ ). See Supplementary Information for details. TBS, *t*-butyldimethylsilyl; Bn, benzyl; PMB, *p*-methoxybenzyl; SEM, 2-(trimethylsilyl)ethoxymethyl.

the enantiomerically pure diol, not easily accessed by an alternative protocol, can serve as precursor to various derivatives. All allene additions proceed with complete  $\alpha$  selectivity (<2% of propargyl products detected).

The ease of accessing the present class of catalysts, the importance of amines and alcohols to the preparation of biologically active molecules, as well as the simplicity, economy and selectivity with which the catalytic transformations proceed, foreshadow a lasting impact on future efforts in catalyst development and chemical synthesis. Development of other efficient and enantioselective C–C bond forming reactions promoted by the present catalyst class is in progress.

## METHODS SUMMARY

**Preparation of catalyst solution.** Aminophenol **2g** (15.0 mg, 0.049 mmol) is weighed out into a 4 ml vial to which is added 263  $\mu$ l of a solution of sodium hydroxide (1.95 mg, 0.049 mmol) in reagent grade methanol (a 111 mg NaOH pellet (Fisher) is dissolved in 15 ml solvent). After removal of solvent, 0.5 ml of technical grade anhydrous toluene is added and concentrated *in vacuo* to remove residual methanol and water. The resulting white solid is dried at 0.5 torr for 30 min and the vial sealed with a cap containing a Teflon septum. Toluene (1.0 ml) is added to yield a suspension.

**Gram-scale procedure for allyl addition.** A round-bottom flask (50 ml, not flame dried, equipped with a magnetic stirring bar) is charged with imine **3a** (1.0 g, 3.3 mmol) and subjected to 0.5 torr for 30 min, purged with dry nitrogen and sealed with a rubber septum. Toluene (30 ml) is added, followed by allylboronic acid pinacol ester **1a** (800  $\mu$ l, 4.26 mmol, 1.3 equiv.) from a septum-sealed bottle (Frontier Scientific, used as received) and methanol (200  $\mu$ l, 4.92 mmol) from a septum-sealed bottle (Acros, 99.9% ExtraDry, used as received). A suspension of the catalyst containing aminophenol **2g** (10.1 mg, 0.033 mmol) and sodium hydroxide (1.31 mg, 0.033 mmol, 0.01 equiv.) in 0.67 ml toluene is added through a syringe to the mixture. After two hours, the solvent is evaporated and the residue is taken up in 30 ml technical grade mixed hexanes. The suspension is subjected to sonication for two minutes, filtered and washed four times with 3 ml hexanes. The product is dried at 0.5 torr and obtained in 92% yield (1.04 g, 3.01 mmol, e.r. 97.5:2.5). Elemental analysis for  $C_{22}H_{22}NOP$ : calculated; C, 76.06; H, 6.38; N, 4.03. Found: C, 75.77; H, 6.43; N 3.98.

Received 28 August; accepted 17 December 2012.

- Lovering, F., Bikker, J. & Humblet, C. Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **52**, 6752–6756 (2009).
- Ojima, I. (ed.) *Catalytic Asymmetric Synthesis* (Wiley, 2010).
- Nakamura, E. & Sato, K. Managing the scarcity of chemical elements. *Nature Mater.* **10**, 158–161 (2011).
- Yus, M., González-Gómez, J. C. & Foubelo, F. Catalytic enantioselective allylation of carbonyl compounds and imines. *Chem. Rev.* **111**, 7774–7854 (2011).
- Borzilleri, R. M. *et al.* A novel application of a Pd(0)-catalyzed nucleophilic substitution reaction to the regio- and stereoselective synthesis of lactam analogues of the epothilone natural products. *J. Am. Chem. Soc.* **122**, 8890–8897 (2000).
- Sirasani, G. & Andrade, R. B. Total synthesis of (–)-leuconicine A and B. *Org. Lett.* **13**, 4736–4737 (2011).
- Xie, W., Zhou, B., Pei, D. & Ma, D. Total synthesis of cyclic tetrapeptide FR235222, a potent immunosuppressant that inhibits mammalian histone deacetylases. *Org. Lett.* **7**, 2775–2777 (2005).
- Kim, S. J. & Jang, D. O. Indium-mediated catalytic enantioselective allylation of *N*-benzoylhydrazones using a protonated chiral amine. *J. Am. Chem. Soc.* **132**, 12168–12169 (2010).
- Tan, K. L. & Jacobsen, E. N. Indium-mediated asymmetric allylation of acylhydrazones using a chiral urea catalyst. *Angew. Chem. Int. Edn* **46**, 1315–1317 (2007).
- Kargbo, R. *et al.* Readily accessible, modular, and tunable BINOL 3,3'-perfluoroalkylsulfones: highly efficient catalysts for enantioselective In-mediated imine allylation. *J. Am. Chem. Soc.* **129**, 3846–3847 (2007).
- Aydin, J., Kumar, K.-s., Sayah, M. J., Wallner, O. A. & Szabó, K. J. Synthesis and catalytic application of chiral 1,1'-bi-2-naphthol and biphenanthrol-based pincer complexes: selective allylation of sulfonylhydrazones with allyl stannane and allyl trifluoroborate. *J. Org. Chem.* **72**, 4689–4697 (2007).
- Wada, R. *et al.* Catalytic enantioselective allylation of ketohymines. *J. Am. Chem. Soc.* **128**, 7687–7691 (2006).
- Lou, S., Moquist, P. N. & Schaus, S. E. Asymmetric allylboration of acyl imines catalyzed by chiral diols. *J. Am. Chem. Soc.* **129**, 15398–15404 (2007).
- Chakrabarti, A., Konishi, H., Yamaguchi, M., Schneider, U. & Kobayashi, S. Indium(I)-catalyzed asymmetric allylation, crotylation, and  $\alpha$ -chloroallylation of hydrazones with rare constitutional and high configurational selectivities. *Angew. Chem. Int. Edn* **49**, 1838–1841 (2010).
- Naodovic, M., Wadamoto, M. & Yamamoto, H. Enantioselective Ag-catalyzed allylation of aldimines. *Eur. J. Org. Chem.* **2009**, 5129–5131 (2009).
- Ferraris, D. *et al.* Catalytic, enantioselective alkylation of  $\alpha$ -imino esters: the synthesis of nonnatural  $\alpha$ -amino acid derivatives. *J. Am. Chem. Soc.* **124**, 67–77 (2002).
- Vieira, E. M., Snapper, M. L. & Hoveyda, A. H. Enantioselective synthesis of homoallylic amines through reactions of (pinacolato)allylborons with aryl-, heteroaryl-, alkyl-, or alkene-substituted aldimines catalyzed by chiral  $C_1$ -symmetric NHC–Cu complexes. *J. Am. Chem. Soc.* **133**, 3332–3335 (2011).
- Hamada, T., Manabe, K. & Kobayashi, S. Catalytic asymmetric allylation of hydrazono esters in aqueous media by using  $ZnF_2$ -chiral diamine. *Angew. Chem. Int. Edn* **42**, 3927–3930 (2003).
- Peddibhotla, S. 3-Substituted-3-hydroxy-2-oxindole, an emerging new scaffold for drug discovery with potential anti-cancer and other biological activities. *Curr. Bioact. Compd.* **5**, 20–38 (2009).
- Coste, A., Couty, F. & Evano, G. TMC-95A–D and analogues: chemistry and biology. *C.R. Chimie* **11**, 1544–1573 (2008).
- Yamamoto, D. *et al.* Design, synthesis, and biological activities of madindoline analogues. *Bioorg. Med. Chem. Lett.* **16**, 2807–2811 (2006).
- Hanhan, N. V., Sahin, A. H., Chang, T. W., Fetting, J. C. & Franz, A. K. Catalytic asymmetric synthesis of 3-hydroxy-2-oxindoles. *Angew. Chem. Int. Edn* **49**, 744–747 (2010).
- Itoh, J., Han, S. B. & Krische, M. J. Enantioselective allylation, crotylation, and reverse prenylation of substituted isatins: iridium-catalyzed C–C bond-forming transfer hydrogenation. *Angew. Chem. Int. Edn* **48**, 6313–6316 (2009).
- Vieira, E. M., Haefner, F., Snapper, M. L. & Hoveyda, A. H. A robust, efficient and highly enantioselective method for synthesis of homopropargyl amines. *Angew. Chem. Int. Edn* **51**, 6618–6621 (2012).
- Barnett, D. S., Moquist, P. N. & Schaus, S. E. The mechanism and an improved asymmetric allylboration of ketones catalyzed by chiral biphenols. *Angew. Chem. Int. Edn* **48**, 8679–8682 (2009).
- Weinreb, S. M. & Orr, R. K. *N*-Phosphinoylimines: an emerging class of reactive intermediates for stereoselective organic synthesis. *Synthesis* **8**, 1205–1227 (2005).
- Fujita, M., Nagano, T., Schneider, U., Hamada, T. & Kobayashi, S. Zn-catalyzed asymmetric allylation for the synthesis of optically active allylglycine derivatives. Regio- and stereoselective formal  $\alpha$ -addition of allylboronates to hydrazono esters. *J. Am. Chem. Soc.* **130**, 2914–2915 (2008).
- Guzman-Martinez, A. & Hoveyda, A. H. Enantioselective synthesis of allylboronates bearing a tertiary or quaternary B-substituted stereogenic carbon by NHC–Cu-catalyzed substitution reactions. *J. Am. Chem. Soc.* **132**, 10634–10637 (2010).
- Itoh, T., Ishikawa, H. & Hayashi, Y. Asymmetric aldol reaction of acetaldehyde and isatin derivatives for the total syntheses of *ent*-convolutamidine E and CPC-1 and a half fragment of madindoline A and B. *Org. Lett.* **11**, 3854–3857 (2009).
- Cravotto, G. *et al.* Convolutamidine A: the first authenticated absolute configuration and enantioselective synthesis. *Tetrahedr. Asym.* **17**, 3070–3074 (2006).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** This research was supported by the US National Institutes of Health, Institute of General Medical Sciences (grant GM-57212). S.T. was a Swiss National Science Foundation Postdoctoral Fellow; E.M.V. was an AstraZeneca Graduate Fellow. We thank B. Li for assistance in securing X-ray structures, S. J. Meek, S. J. Malcolmson and K. L. Tan for discussions, Boston College for providing access to computational facilities and Frontier Scientific for gifts of various organoboron reagents.

**Author Contributions** D.L.S. and T.P. were involved in the discovery, design and development of the catalysts; D.L.S., S.T. and T.P. worked on applications to enantioselective additions to imines; D.L.S. and E.M.V. developed the enantioselective allyl and allene additions to isatins, respectively; D.L.S., S.T., T.P. and F.H. carried out mechanistic and computational studies. A.H.H. conceived, designed and directed the investigations and wrote the manuscript with revisions provided by D.L.S. and E.M.V. This work is part of a collaborative programme between A.H.H. and M.L.S. involving the development of amino acid-derived chiral catalysts.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.H.H. ([amir.hoveyda@bc.edu](mailto:amir.hoveyda@bc.edu)).

# Insolation-induced mid-Brunhes transition in Southern Ocean ventilation and deep-ocean temperature

Qiuzhen Yin<sup>1</sup>

Glacial–interglacial cycles characterized by long cold periods interrupted by short periods of warmth are the dominant feature of Pleistocene climate, with the relative intensity and duration of past and future interglacials being of particular interest for civilization. The interglacials after 430,000 years ago were characterized by warmer climates<sup>1,2</sup> and higher atmospheric concentrations of carbon dioxide<sup>3</sup> than the interglacials before, but the cause of this climatic transition (the so-called mid-Brunhes event (MBE)) is unknown. Here I show, on the basis of model simulations, that in response to insolation changes only, feedbacks between sea ice, temperature, evaporation and salinity caused vigorous pre-MBE Antarctic bottom water formation and Southern Ocean ventilation. My results also show that strong westerlies increased the pre-MBE overturning in the Southern Ocean via an increased latitudinal insolation gradient created by changes in eccentricity during austral winter and by changes in obliquity during austral summer. The stronger bottom water formation led to a cooler deep ocean during the older interglacials. These insolation-induced differences in the deep-sea temperature and in the Southern Ocean ventilation between the more recent interglacials and the older ones were not expected, because there is no straightforward systematic difference in the astronomical parameters between the interglacials before and after 430,000 years ago<sup>4</sup>. Rather than being a real ‘event’, the apparent MBE seems to have resulted from a series of individual interglacial responses—including notable exceptions to the general pattern—to various combinations of insolation conditions. Consequently, assuming no anthropogenic interference, future interglacials may have pre- or post-MBE characteristics without there being a systematic change in forcings. These findings are a first step towards understanding the magnitude change of the interglacial carbon dioxide concentration around 430,000 years ago.

The interglacial climates of the past 1 Myr are characterized by a transition, about 430 kyr ago, between the older ones, which were relatively cool, and the more recent ones, which were relatively warm. This transition corresponds to the so-called MBE<sup>5</sup>. Benthic  $\delta^{18}\text{O}$  records<sup>1</sup> show systematically lower  $\delta^{18}\text{O}$  values, which implies a smaller global ice volume, a higher deep-ocean temperature or both, during the interglacials after 430 kyr ago (post-MBE interglacials, which include Marine Isotope Stages (MISs) 1, 5, 7, 9 and 11) than during those between 800 kyr ago and 430 kyr ago (pre-MBE interglacials, which include MISs 13, 15, 17 and 19). The ice-core records from Antarctica show that, relative to the pre-MBE interglacials, the post-MBE interglacials were warmer<sup>2</sup> (on average by 4 °C) and had atmospheric  $\text{CO}_2$  concentrations about 40 parts per million by volume higher<sup>3</sup>. During the post-MBE interglacials, the surface of the Southern Ocean was much warmer<sup>6,7</sup> and had less sea ice<sup>8</sup>, and the Antarctica polar front was suggested to be located farther to the south<sup>9</sup>. Although the terminology of an event (the MBE) may continue to be

used, my modelling results suggest that there is not necessarily a need to invoke a singular event around 430 kyr ago but rather a transition between two different interglacial regimes.

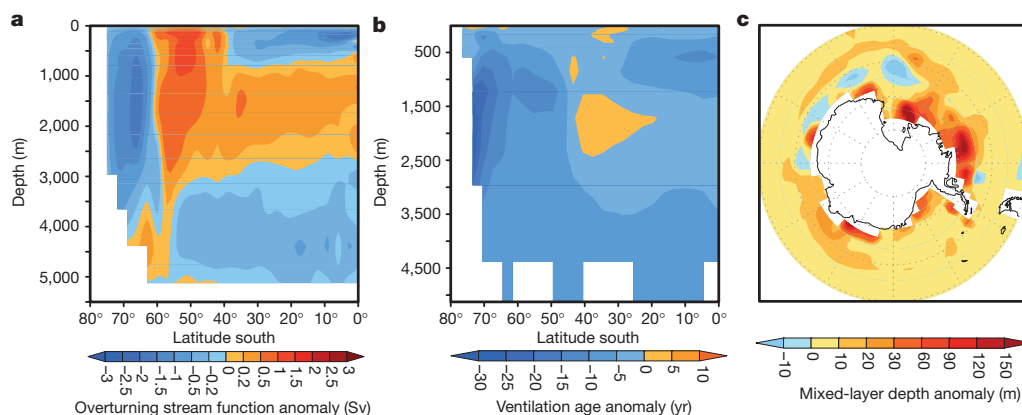
Because insolation is the most important known external forcing of the climate system, it is worthwhile to test whether only insolation change can be at the origin of the systematic differences between the interglacials before and after 430 kyr ago. Here a global atmosphere/ocean/sea-ice/vegetation coupled model (LOVECLIM<sup>10</sup>) is used to investigate the pure impact of insolation on the large-scale oceanic features during all nine interglacials of the past 800 kyr. Two sets of experiments were performed. In the first set, only insolation was allowed to vary between the interglacials, greenhouse gases concentrations (GHGs) being kept constant. This makes it possible to quantify the pure impact of insolation. The second set, in which both insolation and GHGs vary, was used to compare model results to proxy records. This model–proxy comparison is given in Supplementary Information together with details about the model and experiments. Although there are no directly visible systematic differences in the astronomical parameters<sup>4</sup> or in the insolation-induced surface climates<sup>11</sup> between the pre-MBE and the post-MBE interglacials, this study shows that, through a set of internal mechanisms, insolation alone induces a systematic difference between the interglacials before and after 430 kyr ago in some oceanic processes that are critical for the carbon cycle.

In response to insolation only, the differences in the meridional overturning stream function between each of the pre-MBE interglacials (except MIS 19) and each of the post-MBE ones show a similar anomaly structure in the Southern Ocean (Supplementary Fig. 1): the Antarctic bottom water (AABW) cell along the Antarctica shelf and the Deacon cell are both more intense during the pre-MBE interglacials (Fig. 1a). In addition, both ventilation age (Fig. 1b) and mixed-layer depth (Fig. 1c) indicate stronger Southern Ocean ventilation and deeper convection during the pre-MBE interglacials (Supplementary Fig. 2). The consistent behaviour of these three variables indicates that insolation alone generates a systematic difference between the interglacials before and after 430 kyr ago in the Southern Ocean ventilation and in the deep-water formation around Antarctica. There are, however, two noticeable exceptions: MIS 7 in the post-MBE interglacials and MIS 19 in the pre-MBE ones. MIS 7 is indeed a ‘cool’ insolation-induced interglacial and MIS 19 is a ‘warm’ one<sup>11</sup>. These two interglacials are also noticed to be exceptions in the Antarctica ice-core records<sup>2</sup>. Moreover, the magnitude of MIS 7 is similar to that of the pre-MBE interglacials in the benthic  $\delta^{18}\text{O}$  record<sup>1</sup>, and MIS 19 is suggested to be an analogue of our present interglacial<sup>11–13</sup> in respect of the astronomical parameters, the peak  $\text{CO}_2$ -equivalent concentration and the evolution of Antarctica temperature and benthic  $\delta^{18}\text{O}$  records. Therefore, to avoid repetition, the terms ‘pre-MBE’ and ‘post-MBE’ interglacials used hereafter exclude MIS 19 and MIS 7, respectively.

To explain the stronger pre-MBE AABW formation, the density of the upper layer in the Southern Ocean is first analysed. The water

<sup>1</sup>Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, Université Catholique de Louvain, Chemin du Cyclotron 2, 1348 Louvain-la-Neuve, Belgium.





**Figure 1 | Insolation-induced annual mean differences in stream function, ventilation age and mixed-layer depth between the averages of the pre-MBE and post-MBE interglacials.** **a**, Meridional overturning stream function; **b**, ventilation age of the Atlantic; **c**, mixed-layer depth. The two exceptions, MIS 7 and MIS 19, are excluded. The differences in overturning stream function between each individual interglacial are given in Supplementary Fig. 1, and the water age and mixed-layer depth of each interglacial are given in

Supplementary Fig. 2. Compared with the post-MBE interglacials, the pre-MBE ones are, on average, 7% stronger in maximum stream function, 8% younger in ventilation age and 10% larger in mixed-layer depth in the Southern Ocean. For the individual interglacials, the maximum differences in these three variables are respectively 17% (between MIS 15 and MIS 1), 14% and 17% (both between MIS 17 and MIS 1).

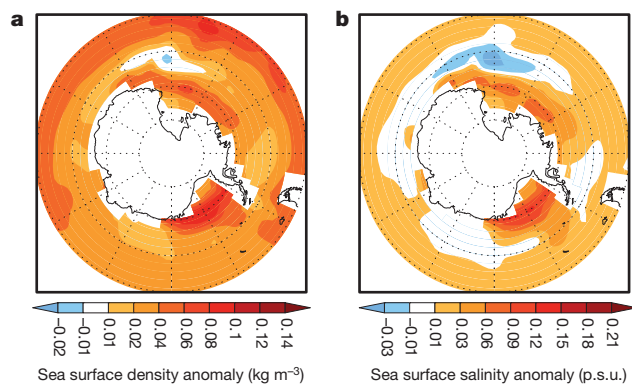
around Antarctica is shown to have a higher surface density (Fig. 2a and Supplementary Fig. 3a) and a weaker stratification during the pre-MBE interglacials. This is particularly true over the marginal seas of western Antarctica extending from the Weddell Sea westwards to the Ross Sea. Being important sites for deep-water formation, the destratification of these seas contributes largely to the intensification of the AABW formation during the pre-MBE interglacials. This larger surface density is primarily a salinity signal (Fig. 2b and Supplementary Fig. 3b), which originates mainly from changes in sea-ice condition and in evaporation.

Compared with the post-MBE interglacials, MIS 13 and MIS 17 have higher sea-ice concentration and thickness around Antarctica (Supplementary Fig. 3c), which, through brine rejection, explains their higher surface salinity. By contrast, MIS 15 has less sea ice. Its higher salinity is attributed to its higher net evaporation (Supplementary Fig. 3d), which results from an increase in evaporation larger than that in precipitation around the coastal sea of Antarctica. Analysing the evaporation of the nine interglacials shows that it is highly correlated with the sea-ice lead fraction and the sea surface temperature. For MIS 15, the high evaporation is due to its increased lead fraction (resulting from reduced sea-ice concentration) and a higher sea-surface temperature. The salting effect of its high evaporation overcomes the freshening effect of its reduced sea ice. By contrast, during

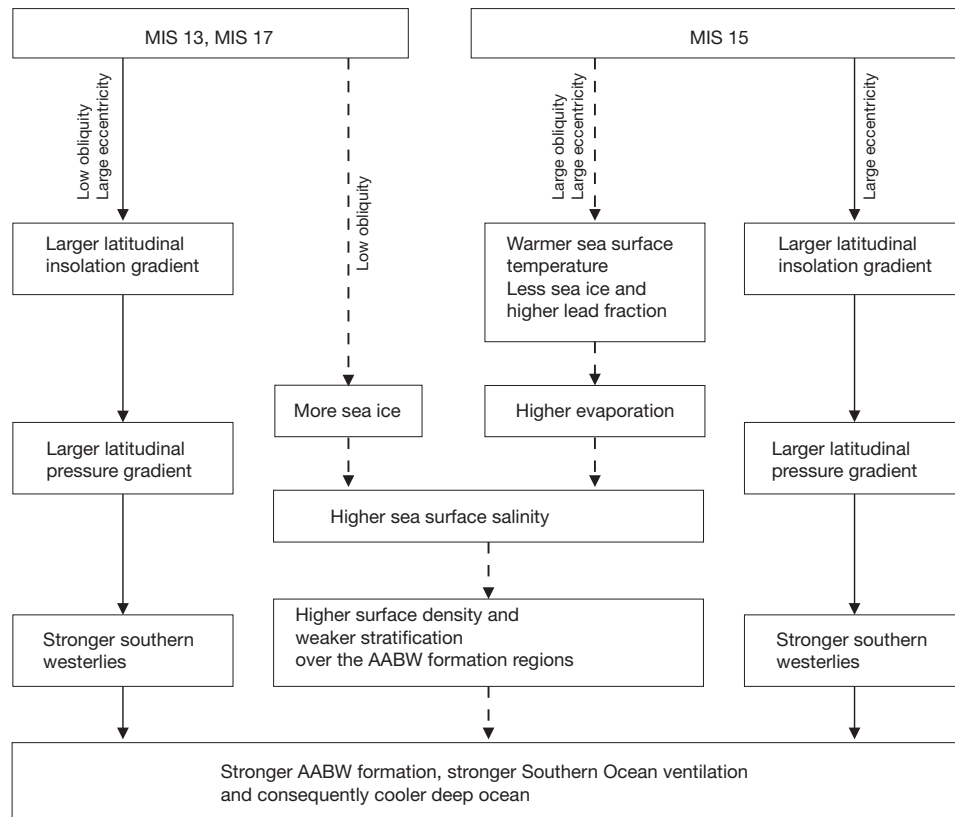
MIS 13 and MIS 17 the salting effect due to more sea ice is more important. However, these different and competing processes lead finally to the same result: MIS 13, MIS 15 and MIS 17 all have higher surface salinity than the post-MBE interglacials. This demonstrates that the internal feedbacks can lead to a systematic difference between the interglacials before and after 430 kyr ago in sea surface salinity, sea surface density and, consequently, AABW formation, although there is no clear systematic difference in the insolation-induced sea surface temperature or sea-ice formation in the Southern Ocean<sup>11</sup>. The schematic representation of the processes and the cause–effect relationship is shown in Fig. 3.

As far as the influence of the astronomical parameters on the southern sea ice is concerned, obliquity has a dominant role through its control on the total local summer solar irradiation, whereas precession (here equivalent to eccentricity) can not be neglected during austral winter<sup>11</sup>. Indeed, larger eccentricity leads to higher summer insolation in the Northern Hemisphere and, consequently, via ocean currents, to more southward oceanic heat transport, which brings heat to around Antarctica and melts sea ice from below. MIS 15 has the second largest obliquity and the largest eccentricity, which explains the minimum sea ice and the highest sea surface temperature over the Southern Ocean during MIS 15. By contrast, more sea ice during MIS 13 and MIS 17 is mainly due to their low obliquity<sup>11</sup> (Supplementary Table 1) which induces a low total summer solar irradiation over the southern high latitudes.

In parallel, insolation induces a stronger annual mean wind stress over the Southern Ocean during the pre-MBE interglacials (Fig. 4). This behaviour of the wind stress is directly linked to the change in the wind and pressure fields. In the Southern Hemisphere, the zonal winds, as well as the latitudinal pressure gradient, are all stronger during the pre-MBE interglacials, with a deepening of the subpolar low and an increase of the subtropical high. Here the seasonal cycle is important because the variation in the annual mean wind stress between the interglacials is mostly due to the change in the wind stress during austral winter (from May to September) and much less during austral summer (from October to April). The difference in the seasonal wind stress between the interglacials can then be explained by the difference between them in the latitudinal gradient of insolation during the respective seasons: the larger the gradient, the stronger the wind stress. During austral winter, the latitudinal gradient of insolation is itself primarily a function (positively correlated) of eccentricity (here equivalent to precession; see Supplementary Information). During austral summer, it is primarily a function (negatively correlated)



**Figure 2 | Insolation-induced annual mean differences in sea-surface density and salinity between the averages of the pre-MBE and post-MBE interglacials.** **a**, Sea surface density; **b**, sea surface salinity. The two exceptions, MIS 7 and MIS 19, are excluded. The differences between each individual interglacial are given in Supplementary Fig. 3. p.s.u., practical salinity units.

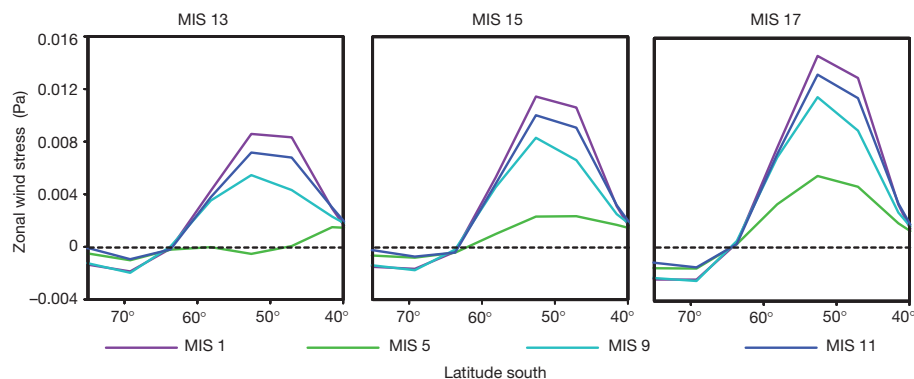


**Figure 3 | Schematic representation of astronomically induced responses in MISs 13, 15 and 17 as compared with the post-MBE interglacials.** Thermohaline-driven changes are indicated by dashed arrows and wind-driven ones by solid arrows.

of obliquity (Supplementary Fig. 6). Consequently, the variance of the wind stress during austral winter is mostly explained by eccentricity, and obliquity is important in austral summer. The annual mean wind stress is therefore finally highly and positively correlated with eccentricity, and to a much lesser degree negatively correlated with obliquity. It is maximum for large eccentricity and low obliquity, as is the case during MIS 13 and MIS 17 (Fig. 3). At MIS 15, the same conclusion holds because its large obliquity only moderates the effect of its large eccentricity. These finally lead to an annual mean wind stress larger (by 5% on average) during the pre-MBE interglacials than during the post-MBE ones. The stronger westerlies during the pre-MBE interglacials are responsible for the stronger upwelling in the Southern Ocean. According to some sensitivity studies<sup>14,15</sup>, they should also contribute to the intensification of the AABW formation.

The intensification of the AABW formation cools down the deep ocean during the pre-MBE interglacials (Supplementary Fig. 4),

although this cooling might be underestimated by the model (Supplementary Information). In addition, the simulated surface cooling over the whole Southern Ocean during MIS 13 and MIS 17 further contributes to this deep-ocean cooling. It is worth noting that the insolation-induced deep-sea cooling is amplified when  $\text{CO}_2$  is allowed to vary. Indeed, when, additionally, the  $\text{CO}_2$  forcing is taken into account, the lower  $\text{CO}_2$  during the pre-MBE interglacials reinforces the cooling caused by insolation alone, leading to a cooler deep ocean during the pre-MBE interglacials than during the post-MBE ones (Supplementary Information). This implies that the lower benthic  $\delta^{18}\text{O}$  magnitude of pre-MBE interglacials, which is a function of both global ice volume and deep-sea temperature, results at least partly from these interglacials' cooler deep-sea temperature and does not necessarily indicate larger continental ice sheets. This is in line with the implications based on some proxy records<sup>16–18</sup> and other modelling results<sup>19</sup>.



**Figure 4 | Insolation-induced annual mean differences in zonal wind stress between each of the pre-MBE interglacials and each of the post-MBE ones.** The positive values north of latitude  $65^\circ\text{S}$  and the negative ones south of  $65^\circ\text{S}$

indicate stronger westerly and, respectively, easterly wind stress. The largest difference in the zonal wind stress is 8% (between MIS 17 and MIS 1).

The simulations discussed here show the a priori unexpected role of the astronomically induced incoming solar radiation in generating a systematic difference between the interglacials before and after 430 kyr ago in the Southern Ocean and in the deep sea. They demonstrate the importance of sea ice, evaporation, salinity and wind, whose complex interactions here shape the final response of the climate system. Because the effect of insolation had to be separated from the effect of CO<sub>2</sub>, these findings could only be discovered through numerical simulations taking into account the atmosphere–ocean coupling and singularizing the pure impact of insolation. My modelling results demonstrate that the simulated systematic differences between the interglacials before and after 430 kyr ago result solely from individual responses to various combinations of the astronomical parameters, and therefore that it is not necessary to invoke a singular event around 430 kyr ago to explain the transition. The reasonable agreement between model results and proxy records (Supplementary Information) supports both the present simulations and the proxy-based reconstructions. Because the Southern Ocean and the deep sea are critical in the carbon cycle<sup>15,20–22</sup>, finding insolation-induced systematic differences in these components between the interglacials before and after 430 kyr ago is a first step towards understanding the origin of the magnitude change of the interglacial CO<sub>2</sub> concentration around 430 kyr ago. Coupled climate/carbon-cycle models appropriate for long-term integration are needed to confirm the triggering role of insolation in the transition between the two interglacial regimes.

Received 28 July; accepted 13 November 2012.

- Lisiecki, L. E. & Raymo, M. E. A. Pliocene–Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* **20**, PA1003 (2005).
- Jouzel, J. *et al.* Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science* **317**, 793–796 (2007).
- Lüthi, D. *et al.* High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* **453**, 379–382 (2008).
- Yin, Q. Z. & Berger, A. Insolation and CO<sub>2</sub> contribution to the interglacial climate before and after the mid-Brunhes event. *Nature Geosci.* **3**, 243–246 (2010).
- Jansen, J. H. F., Kuijpers, A. & Troelstra, S. R. A. Mid-Brunhes climatic event: long-term changes in global atmosphere and ocean circulation. *Science* **232**, 619–622 (1986).
- Becquey, S. & Gersonde, R. Past hydrographic and climatic changes in the Subantarctic Zone of the South Atlantic – the Pleistocene record from ODP Site 1090. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **182**, 221–239 (2002).
- Schaefer, G. *et al.* Planktic foraminiferal and sea surface temperature record during the last 1 Myr across the Subtropical Front, Southwest Pacific. *Mar. Micropaleontol.* **54**, 191–212 (2005).
- Wolff, E. W. *et al.* Southern Ocean sea-ice extent, productivity and iron flux over the past eight glacial cycles. *Nature* **440**, 491–496 (2006).
- Kemp, A. E. S., Grigorov, I., Pearce, R. B. & Naveira Garabato, A. C. Migration of the Antarctic Polar Front through the mid-Pleistocene transition: evidence and climatic implications. *Quat. Sci. Rev.* **29**, 1993–2009 (2010).
- Goosse, H. *et al.* Description of the Earth system model of intermediate complexity LOVECLIM version 1.2. *Geosci. Model Dev.* **3**, 603–633 (2010).
- Yin, Q. Z. & Berger, A. Individual contribution of insolation and CO<sub>2</sub> to the interglacial climates of the past 800,000 years. *Clim. Dyn.* **38**, 709–724 (2012).
- Berger, A., & Loutre, M. F. Modeling the climate response to the astronomical and CO<sub>2</sub> forcings. *C. R. Acad. Sci. III* **323**, 1–16 (1996).
- Tzedakis, P. C. The MIS 11 - MIS 1 analogy, southern European vegetation, atmospheric methane and the “early anthropogenic hypothesis”. *Clim. Past* **6**, 131–144 (2010).
- Rahmstorf, S. & England, M. H. Influence of Southern Hemisphere winds on North Atlantic deep water flow. *J. Phys. Oceanogr.* **27**, 2040–2054 (1997).
- Menviel, L., Timmermann, A., Mouchet, A. & Timm, O. Climate and marine carbon cycle response to changes in the strength of the Southern Hemispheric westerlies. *Paleoceanography* **23**, PA4201 (2008).
- Elderfield, H. *et al.* Evolution of ocean temperature and ice volume through the mid-Pleistocene climate transition. *Science* **337**, 704–709 (2012).
- Siddall, M., Honisch, B., Waelbroeck, C. & Huybers, P. Changes in deep Pacific temperature during the mid-Pleistocene transition and Quaternary. *Quat. Sci. Rev.* **29**, 170–181 (2010).
- Guo, Z. T., Berger, A., Yin, Q. Z. & Qin, L. Strong asymmetry of hemispheric climates during MIS-13 inferred from correlating China loess and Antarctica ice records. *Clim. Past* **5**, 21–31 (2009).
- Ganopolski, A. & Calov, R. The role of orbital forcing, carbon dioxide and regolith in 100 kyr glacial cycles. *Clim. Past* **7**, 1415–1425 (2011).
- Skinner, L. C., Fallon, S., Waelbroeck, C., Michel, E. & Barker, S. Ventilation of the deep Southern Ocean and deglacial CO<sub>2</sub> rise. *Science* **328**, 1147–1151 (2010).
- Martin, P., Archer, D. & Lea, D. Role of deep sea temperatures in the carbon cycle during the Last Glacial. *Paleoceanography* **20**, PA2015 (2005).
- Stott, L., Timmermann, A. & Thunell, R. Southern Hemisphere and deep-sea warming led deglacial atmospheric CO<sub>2</sub> rise and tropical warming. *Science* **318**, 435–438 (2007).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** Thanks to A. Berger, M. Crucifix, A. Ganopolski and D. Paillard for their comments on the previous draft of this paper, to A. Mouchet for her help on the water age calculation in LOVECLIM, and to A. Timmermann for his comments on the results. Thanks also to N. Herold for help with English. This work is supported by the European Research Council Advanced Grant EMIS (no. 227348 of the Programme ‘Ideas’). The author is supported by the Belgian National Fund for Scientific Research (FRS-FNRS). Access to computer facilities was made easier through sponsorship from S. A. Electrabel, Belgium.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The author declares no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to the author ([qiuzhen.yin@uclouvain.be](mailto:qiuzhen.yin@uclouvain.be)).



# Vertebral architecture in the earliest stem tetrapods

Stephanie E. Pierce<sup>1,2</sup>, Per E. Ahlberg<sup>3</sup>, John R. Hutchinson<sup>2</sup>, Julia L. Molnar<sup>2</sup>, Sophie Sanchez<sup>3,4</sup>, Paul Tafforeau<sup>4</sup> & Jennifer A. Clack<sup>1</sup>

The construction of the vertebral column has been used as a key anatomical character in defining and diagnosing early tetrapod groups<sup>1</sup>. Rhachitinous vertebrae<sup>2</sup>—in which there is a dorsally placed neural arch and spine, an anteroventrally placed intercentrum and paired, posterodorsally placed pleurocentra—have long been considered the ancestral morphology for tetrapods<sup>1,3–6</sup>. Nonetheless, very little is known about vertebral anatomy in the earliest stem tetrapods, because most specimens remain trapped in surrounding matrix, obscuring important anatomical features<sup>7–9</sup>. Here we describe the three-dimensional vertebral architecture of the Late Devonian stem tetrapod *Ichthyostega* using propagation phase-contrast X-ray synchrotron microtomography. Our scans reveal a diverse array of new morphological, and associated developmental and functional, characteristics, including a possible posterior-to-anterior vertebral ossification sequence and the first evolutionary appearance of ossified sternal elements. One of the most intriguing features relates to the positional relationships between the vertebral elements, with the pleurocentra being unexpectedly sutured or fused to the intercentra that directly succeed them, indicating a ‘reverse’ rhachitinous design<sup>10</sup>. Comparison of *Ichthyostega* with two other stem tetrapods, *Acanthostega*<sup>7</sup> and *Pederpes*<sup>8</sup>, shows that reverse rhachitinous vertebrae may be the ancestral condition for limbed vertebrates. This study fundamentally revises our current understanding<sup>11</sup> of vertebral column evolution in the earliest tetrapods and raises questions about the presumed vertebral architecture of tetrapodomorph fish<sup>12,13</sup> and later, more crownward, tetrapods.

Jarvik<sup>3,14,15</sup> reassembled the vertebral column of the Devonian stem tetrapod *Ichthyostega* from ‘imperfectly’ preserved material. Each vertebral unit was thought to consist of a relatively straight neural spine; paired, round pleurocentra situated underneath the neural arch in front and spanning the intervertebral joint; and an anteroventrally placed crescent-shaped intercentrum (Fig. 1a, b). The morphology and positional arrangement of the vertebral elements in *Ichthyostega* was primarily fashioned after Jarvik’s restoration of the vertebrae in *Eusthenopteron*<sup>3</sup>—a tetrapodomorph fish considered to possess tetrapod-like rhachitinous vertebrae<sup>1</sup>. Extensive re-examination of the *Ichthyostega* fossil material revealed pronounced regionalization of the neural arch/spine morphology, and resulted in a reconstruction of the vertebral column differing radically from Jarvik’s interpretation<sup>9</sup>. However, little information could be gleaned about the remaining vertebral elements as they were concealed by other anatomical features (for example, ribs) or remained embedded in matrix.

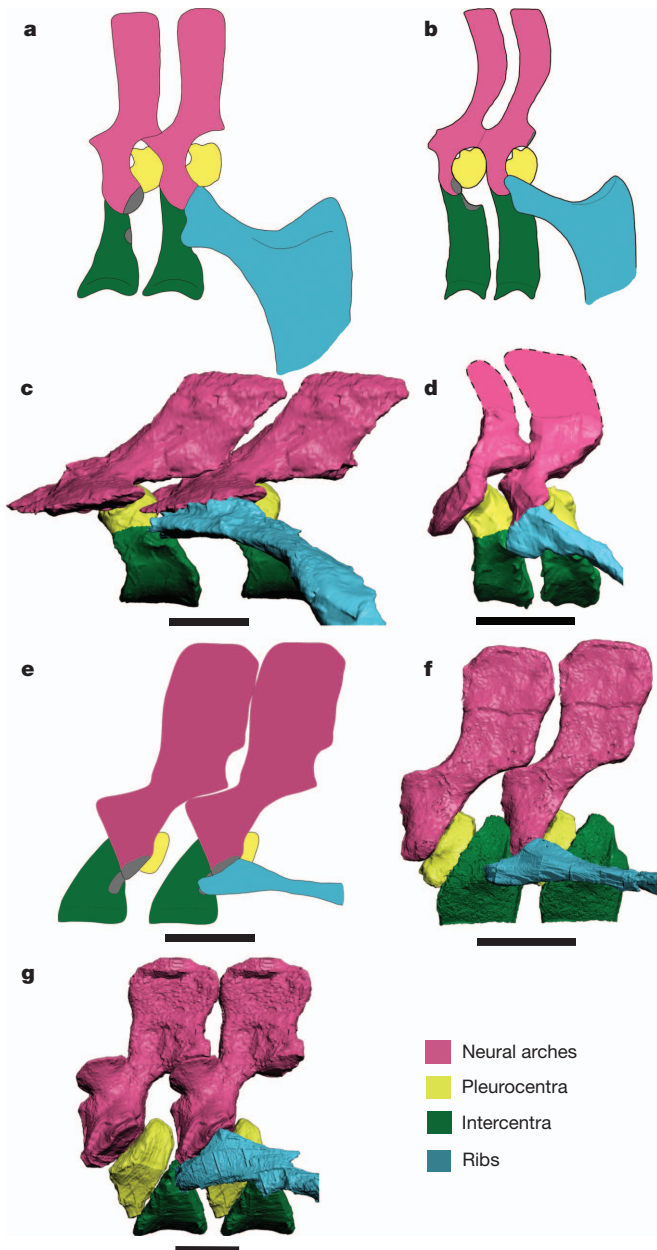
Recently, the three-dimensional skeletal anatomy of *Ichthyostega* was recreated from microcomputed tomography ( $\mu$ CT) scans<sup>16</sup>. The  $\mu$ CT data exposed all elements within the vertebral column and indicated the presence of new anatomical features, as well as an unexpected vertebral configuration. In light of this, two key specimens were subjected to phase-contrast X-ray synchrotron microtomography (PPC-SR $\mu$ CT) using a specially developed scanning protocol<sup>17</sup> (see Methods) to obtain the highest possible resolution and sensitivity to

low-contrast structures. The specimens consist of an almost complete thoracic region with associated ribs (MGUH VP 6115) and a ‘lumbar’ region preserving most elements except for the more posterior neural arches/spines (MGUH VP 29017a). For comparison, a section of the vertebral column from the Devonian tetrapod *Acanthostega gunnari* (MGUH f.n. 1227) was also scanned using PPC-SR $\mu$ CT, and the skeletal remains of the more crownward Tournaisian tetrapod *Pederpes finneyae* (GLAHMS 100815) were imaged using  $\mu$ CT. For visualization of the segmented scan data see Supplementary Videos 1–4.

As has been previously noted<sup>9</sup>, most of the thoracic neural spines of *Ichthyostega* are caudally inclined (Figs 1c and 2); however, our PPC-SR $\mu$ CT data have revealed that the anterior thoracic neural spines, flanked by the cleithra, are vertical in orientation (Fig. 2c, d). In addition, we have discovered further regionalization in transverse process morphology (Figs 1c, d, 2 and 3). The anterior thoracic and lumbar vertebral arches have ventrally projecting transverse processes, whereas those in the posterior thoracic region are horizontally flat, and square-shaped, with small triangular-shaped anterior extensions (see further below). Rib heads in the thoracic region are not bicipital (or incipiently bicipital; Fig. 1e, f, g), but rather anteroposteriorly elongated—articulating with the entire length of the transverse processes (Figs 1c and 2c, e). Comparing vertebral arch morphology with that of *Acanthostega*<sup>7</sup> (Fig. 1e, f) and *Pederpes*<sup>8</sup> (Fig. 1g), we suggest that vertically oriented neural spines and ventrally projecting transverse processes are the general condition for the tetrapod stem group. This indicates that the posterior thorax, rather than the lumbar region<sup>9</sup>, is uniquely constructed in *Ichthyostega*. Considering that vertebral processes tend to align themselves perpendicular to the predominant forces acting on them<sup>18,19</sup>, the morphology of the posterior thoracic vertebrae in *Ichthyostega* signifies a departure from the primary role of the epaxial musculature.

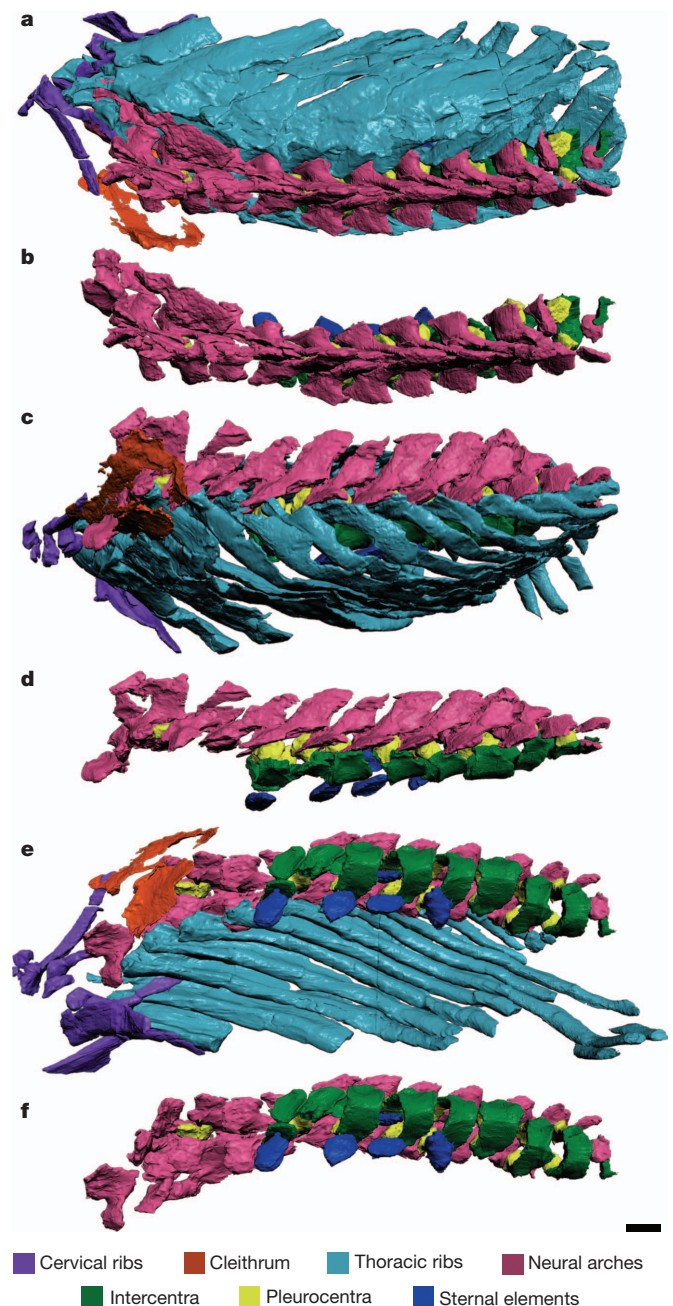
Except for a possible pair of small pleurocentra sitting between the third and fourth neural arches (Fig. 2d–f), there is a complete lack of ossified centrum elements in the anterior part of the thoracic column (Figs 2 and 3). Fully developed intercentrum/pleurocentra units first appear around the fifth preserved neural arch. The articulated condition of the specimen (Supplementary Fig. 1 and Supplementary Video 1), including the skull preserved in anatomically correct position<sup>9,15</sup>, suggests that this pattern is not a preservational artefact and may indicate an unusual posterior-to-anterior ossification sequence. Although vertebral development most commonly occurs in an anterior-to-posterior direction<sup>20,21</sup> (possibly linked to somitogenesis<sup>22</sup> and *Hox* gene expression patterns<sup>23</sup>), some fish have been shown to exhibit the reverse pattern (for example, *Eusthenopteron* and various actinopterygians)<sup>21</sup>. As these fish tend to be phylogenetically disparate, it has been suggested that a posterior-to-anterior ossification sequence may be related to functional constraints, such as the locomotor behaviours of hatchlings and juveniles<sup>21</sup>. Such a hypothesis is consistent with *Ichthyostega* being primarily aquatic in habit as a juvenile<sup>24</sup> and using its paddle-like hindlimbs and tail for swimming<sup>16</sup>, although alternative explanations may exist.

<sup>1</sup>University Museum of Zoology, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. <sup>2</sup>Department of Comparative Biomedical Sciences and Structure & Motion Laboratory, The Royal Veterinary College, Hawkshead Lane, Hatfield AL9 7TA, UK. <sup>3</sup>Subdepartment of Evolution and Development, Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18A, 752 36 Uppsala, Sweden. <sup>4</sup>European Synchrotron Radiation Facility, 6 Rue Jules Horowitz, BP 220, 38043 Grenoble Cedex, France.



**Figure 1 | Virtual restoration of vertebral elements in the earliest stem tetrapods, depicted in left lateral view.** **a**, *Ichthyostega* dorsal vertebrae taken from ref. 14. **b**, *Ichthyostega* dorsal vertebrae taken from ref. 15. **c**, *Ichthyostega* posterior thoracic vertebrae recreated from PPC-SRμCT data (MGUH VP 6115). **d**, *Ichthyostega* anterior lumbar vertebrae recreated from PPC-SRμCT data (MGUH VP 29017a; neural spine tips after MGUH VP 6054 and ref. 9). **e**, *Acanthostega* dorsal vertebrae taken from ref. 7. **f**, *Acanthostega* anterior dorsal vertebrae recreated from PPC-SRμCT data (MGUH f.n. 1227). **g**, *Pederpes* posterior dorsal vertebrae recreated from μCT data (GLAHMS 100815). All vertebrae scaled to the same dorsoventral height. Scale bars, 10 mm.

The centra of *Ichthyostega* display a remarkable array of morphological features. In the thoracic region, the intercentra are long anteroposteriorly and widely spaced (Figs 1c and 2d–f); however, they become noticeably shorter and more closely packed in the lumbar region (Figs 1d and 3a–c). As the dimensions of vertebral bodies have an effect on passive flexibility at intervertebral joints<sup>25</sup>, shortening of the lumbar intercentra (in combination with changes in neural arch morphology) indicate a functional shift in this region. In particular, short, closely packed centra indicate increased intervertebral joint stiffness in the lumbar region due to a minimal degree of angular deflection before adjacent vertebrae obstruct each other<sup>18</sup>. Furthermore, all the

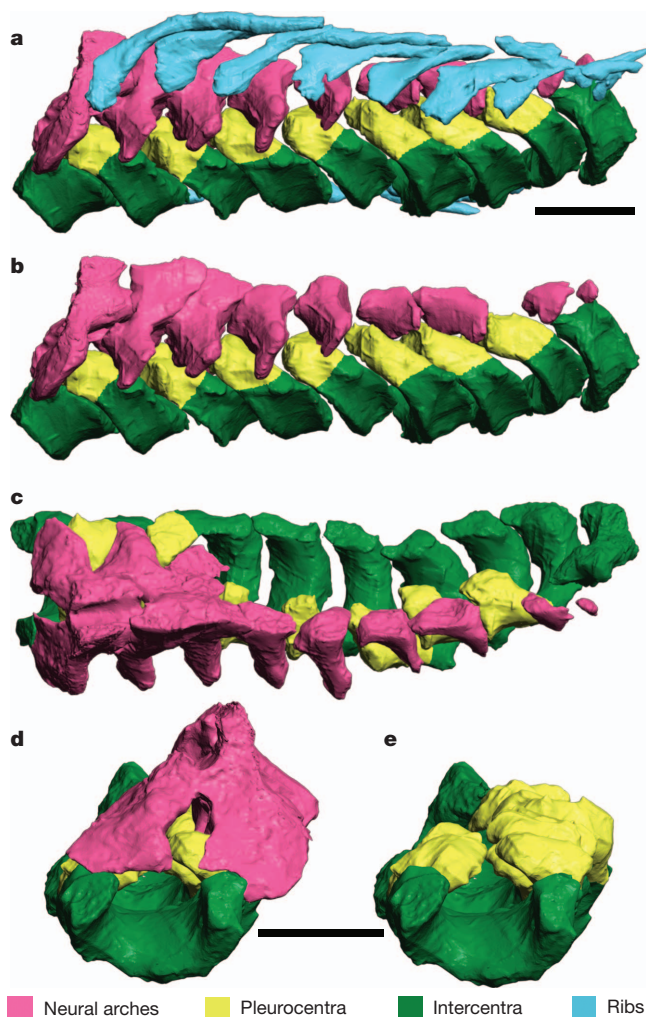


**Figure 2 | Thoracic region of *Ichthyostega* (MGUH VP 6115) rendered from PPC-SRμCT data.** **a**, **b**, Dorsal view of whole specimen (**a**) and vertebral elements only (**b**). **c**, **d**, Lateral view of whole specimen (**c**) and vertebral elements only (**d**). **e**, **f**, Ventral view of whole specimen with left ribs removed (**e**) and vertebral elements only (**f**). The ribcage has been mediolaterally compressed, entombing the vertebral column. On the basis of intervertebral spacing, the first preserved centrum has probably been displaced posteriorly and should correspond to the preceding neural arch. In turn, the second preserved centrum and associated sternal element are disarticulated and sitting inside the succeeding intercentrum. The boundary between fused pleurocentra/intercentra is approximated. Anterior towards left. Scale bar, 10 mm.

left and right intercentra halves appear fused along the length of the vertebral column, except for the first two which are distinctly paired (Figs 2d–f and 3 and Supplementary Fig. 1). The delayed fusion of the anterior intercentra lends further support to a posterior-to-anterior ossification sequence, as intercentra are thought to ossify by two condensations lateral to the notochord before fusing along the midline<sup>26</sup>.

Directly below the first five preserved thoracic intercentra is a series of mid-ventral circular discs. These elements are evenly spaced, one for





**Figure 3 | Lumbar region of *Ichthyostega* (MGUH VP 29017a) rendered from PPC-SRμCT data. a, b, Lateral view of whole specimen (a) and vertebral elements only (b). c, Dorsal view of vertebral elements. d, e, Anterior view of vertebral elements (d) and centrum elements only (e). Most neural spines, right transverse processes and pleurocentra are not preserved; centrum elements are tipped anteriorly; and rib ends are displaced dorsally. The first preserved transverse processes seem to be transitional between the horizontally flat transverse processes of the thoracic region and the smaller and more vertically oriented ones in the remainder of the lumbar region. The boundary between the fused pleurocentra/intercentra is approximated. Anterior towards left. Scale bar, 10 mm.**

each centrum unit (in blue, Fig. 2b, d–f). The PPC-SRμCT data show that the micro-anatomy of these elements is very similar to the other endochondral bones in the specimen (Supplementary Fig. 2). The consistent shape, segmental arrangement, and probable endochondral origin indicate that these mid-ventral elements may represent ossifications (sternebrae) within a cartilaginous sternum, which has been displaced into the dorsal part of the body cavity along with an ulna after death (Supplementary Fig. 3). Whether or not the ribs articulated with the sternebrae cannot be determined, as the distal ends of the anterior ribs are missing. However, the edges of the sternebrae show no noticeable pitting for reception of the rib ends. In living tetrapods, the sternum (or sternebrae) are thought to strengthen the body wall and aid in locomotion and, in some amniotes, ventilation<sup>27</sup>. Conceivably, the presence of sternebrae in the anterior thorax of *Ichthyostega* may have helped to reinforce the ribcage during synchronous forelimb ‘crutching’ on land<sup>16</sup>, allowing the animal to balance its body weight on its chest during the swing phase of forelimb movement.

The pleurocentra of *Ichthyostega* are positioned posteroventrally to their respective neural arch throughout the column (Figs 1c, d, 2d and 3); however, in the thoracic region they are clasped laterally by triangular extensions of the following transverse processes (Figs 1c and 2d). This anatomical arrangement essentially forms an intervertebral locking mechanism in the thorax, restricting the range of joint mobility. A more evolutionarily important feature of the vertebral column is that the pleurocentra are either sutured (anteriorly) or fused (posteriorly) to the anterodorsal aspect of the intercentrum that directly succeed them (Figs 1c, d, 2d and 3 and Supplementary Videos 5 and 6). This indicates a ‘reverse’ rhachitomous condition<sup>10</sup>, with one vertebra encompassing a dorsally placed neural arch above and a posteroventrally placed pleurocentra–intercentrum unit (Fig. 1c, d). On the basis of this discovery, our re-evaluation of the articulated vertebrae in *Acanthostega*<sup>7</sup> (Fig. 1e and Supplementary Video 3) shows that this stem tetrapod also possesses reverse rhachitomous vertebrae, and that *Pederpes*<sup>8</sup> can be interpreted in this way (Fig. 1g and Supplementary Fig. 4). However, rather than being fused to the dorsal aspect as in *Ichthyostega*, the pleurocentra of these taxa articulate at a 45° angle with the anterior border of the intercentrum (Supplementary Fig. 5).

A similar anterior pleurocentra association (or ‘anteropleural’) has, in the past, also been proposed for an assortment of ‘osteolepids’ and even some temnospondyls<sup>10</sup>, although the traditional arrangement is usually attributed to the latter. The potentially widespread distribution of the reverse rhachitomous condition may indicate either a misinterpretation of the fossil record or that the multipartite vertebral structure was developmentally plastic in stem tetrapods. In living tetrapods, vertebrae form by a lineage-restricted process called resegmentation, in which the posterior and anterior sclerotomal halves of consecutive somites fuse<sup>28</sup>. Conversely, the vertebrae of fish develop through notochordal signalling<sup>29</sup> and leaky resegmentation<sup>30</sup>, such that cells cross anterior–posterior domains. This more variable vertebral patterning mechanism may explain why both normal and reverse rhachitomous vertebrae appear in the fossil record, and in some cases, why both conditions have been recognized in closely related species<sup>10</sup>. Perhaps the plasticity of tetrapod vertebral morphology stabilized over the course of evolution, with modern forms acquiring a conserved, unit-partite vertebral structure.

The vertebral architecture of *Ichthyostega* described here is markedly different from Jarvik’s textbook reconstruction<sup>3,14,15</sup> (Fig. 1a, b). It displays even more regionalization than previously proposed<sup>9</sup>, an apparent posterior-to-anterior ossification sequence, and various functional innovations—including the development of sternebrae. Furthermore, the individual vertebral units reveal an unambiguous reverse rhachitomous design, with pleurocentra fused to each succeeding intercentrum (Figs 1c, d, 2 and 3). This centrum arrangement is also present in *Acanthostega* and possibly *Pederpes* (Fig. 1f, g), suggesting that the reverse rhachitomous arrangement could be the ancestral tetrapod condition. Overall, the findings described here demonstrate that the traditional rhachitomous model<sup>1,2</sup> has been misguidedly projected onto the earliest limbed vertebrates<sup>3,7,8</sup>, producing inaccurate reconstructions that have been replicated in the literature for more than half a century<sup>4–6,11,14,15,27</sup>. This study raises questions about the interpretation of vertebral anatomy in tetrapodomorph fish, such as *Eusthenopteron*<sup>12</sup> and *Panderichthys*<sup>13</sup>, and more crownward tetrapods, and has broad ramifications for our future understanding of early tetrapod skeletal evolution.

## METHODS SUMMARY

PPC-SRμCT scans were performed at the European Synchrotron Radiation Facility (ID19, France). They were scanned using a white beam, filtered with 3–9 mm of copper and 60 mm of aluminium<sup>17</sup>, and the W150 wiggler (gap: 38–80 mm), thereby leading to energy with a narrow bandwidth ( $83.7\text{--}133.2 \pm 15$  keV full-width at half-maximum). Samples were imaged through a single crystal LuAG:Ce scintillator of 750 μm thick, coupled with a CCD FReLoN camera placed at 4–9 m from the sample (voxel size: 29.4 μm) (see Methods for details). The scan data were scaled



isotropically to 58.8 µm for segmentation. The µCT scans were performed on an X-Tek HMX 225KeV system, at a voxel size of 90 µm using 175 kV, 127 µA and a 0.5 mm copper filter. All scan data were segmented in Mimics software (Materialise) and rendered in Autodesk 3D Studio Max. *Ichthyostega* and *Acanthostega* specimens are housed in the Geological Museum, Copenhagen (MGUH), and *Pederpes* in the Hunterian Museum, Glasgow (GLAHMS).

**Full Methods** and any associated references are available in the online version of the paper.

**Received 17 August; accepted 7 December 2012.**

**Published online 13 January 2013.**

- Romer, A. S. Review of the Labyrinthodontia. *Bull. Mus. Comp. Zool. Harv.* **99**, 368 (1947).
- Cope, E. D. Geology and palaeontology. *Am. Nat.* **5**, 324–328 (1878).
- Jarvik, E. On the fish-like tail in the ichthyostegid stegocephalians. *Medd. Gronl.* **114**, 1–90 (1952).
- Romer, A. S. The skeleton of the lower Carboniferous labyrinthodont *Pholidogaster pisciformis*. *Bull. Mus. Comp. Zool. Harv.* **131**, 129–159 (1964).
- Panchen, A. L. The homologies of the labyrinthodont centrum. *Evolution* **21**, 24–33 (1967).
- Panchen, A. L. in *Problems in Vertebrate Evolution* (ed. Andrews, S. & Miles, R. S.) 289–318 (Academic, 1977).
- Coates, M. I. The Devonian tetrapod *Acanthostega gunnari* Jarvik: postcranial anatomy, basal tetrapod interrelationships and patterns of skeletal evolution. *Trans. R. Soc. Edinb. Earth Sci.* **87**, 363–421 (1996).
- Clack, J. A. An early tetrapod from 'Romer's Gap'. *Nature* **418**, 72–76 (2002).
- Ahlberg, P. E., Clack, J. A. & Blom, H. The axial skeleton of the Devonian tetrapod *Ichthyostega*. *Nature* **437**, 137–140 (2005).
- Shishkin, M. A. The axial skeleton of early amphibians and the origin of resegmentation in tetrapod vertebrae. *Prog. Zool.* **35**, 180–195 (1989).
- Benton, M. J. *Vertebrate Palaeontology* 3rd edn (Blackwell, 2005).
- Andrews, S. M. & Westoll, T. S. The postcranial skeleton of *Eusthenopteron foordi* Whiteaves. *Trans. R. Soc. Edinb.* **68**, 207–329 (1970).
- Vorobyeva, E. I. & Schultze, H.-P. in *Origins of the Higher Groups of Tetrapods: Controversy and Consensus* (eds Schultze, H.-P. & Trueb, L.) 68–109 (Cornell Univ. Press, 1991).
- Slijper, E. *Comparative Biologic-Anatomical Investigations on the Vertebral Column and Spinal Musculature of Mammals* (North-Holland, 1946).
- Pierce, S. E., Clack, J. A. & Hutchinson, J. R. Comparative axial morphology in pinnipeds and its correlation with aquatic locomotory behaviour. *J. Anat.* **219**, 502–514 (2011).
- Carroll, R. L., Kuntz, A. & Albright, K. Vertebral development and amphibian evolution. *Evol. Dev.* **1**, 36–48 (1999).
- Cote, S., Carroll, R., Cloutier, R. & Bar-Sagi, L. Vertebral development in the Devonian sarcopterygian fish *Eusthenopteron foordi* and the polarity of vertebral evolution in non-amniote tetrapods. *J. Vertebr. Paleontol.* **22**, 487–502 (2002).
- Dequéant, M.-L. & Pourquié, O. Segmental patterning of the vertebrate embryonic axis. *Nature Rev. Genet.* **9**, 370–382 (2008).
- Wellik, D. M. *Hox* patterning of the vertebrate axial skeleton. *Dev. Dyn.* **236**, 2454–2463 (2007).
- Callier, V., Clack, J. A. & Ahlberg, P. E. Contrasting developmental trajectories in the earliest known tetrapod forelimbs. *Science* **324**, 364–367 (2009).
- Long, J. H., Pabst, D. A., Shepherd, W. R. & McLellan, W. A. Locomotor design of dolphin vertebral columns: bending mechanics and morphology of *Delphinus delphis*. *J. Exp. Biol.* **200**, 65–81 (1997).
- Milner, A. R. & Sequeira, S. E. K. The temnospondyl amphibians from the Viséan of East Kirkton, West Lothian, Scotland. *Trans. Roy. Soc. Edin.* **84**, 331–361 (1993).
- Hildebrand, M. & Goslow, G. *Analysis of Vertebrate Structure* 5th edn (John Wiley & Sons, 2001).
- Aoyama, H. & Asamoto, K. The developmental fate of rostral/caudal half of a somite for vertebra and rib formation: experimental confirmation of the resegmentation theory using chick-quail chimeras. *Mech. Dev.* **99**, 71–82 (2000).
- Fleming, A., Keynes, R. & Tannahill, D. A central role for the notochord in vertebral patterning. *Development* **131**, 873–880 (2004).
- Morin-Kensicki, E. M., Melancon, E. & Eisen, J. S. Segmental relationship between somites and vertebral column in zebrafish. *Development* **129**, 3851–3860 (2002).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, for access to beamline ID19; S. Schmidt for support using the µCT scanner at Abertay University; J. Liston and the Glasgow Hunterian Museum for access to fossils in their care; and G. Cuny for access to collections housed in the Geological Museum at the University of Copenhagen. This research was supported by NERC grants NE/G005877/1 and NE/G00711X/1 (J.A.C., J.R.H. and S.E.P.) and ERC grant 233111 (P.E.A. and S.S.).

**Author Contributions** The second to fifth authors have been arranged in alphabetical order. S.E.P., P.E.A., J.R.H. and J.A.C. conceived and designed the project. S.E.P., P.E.A. and J.A.C. analysed and interpreted the data. S.E.P. developed the manuscript, including the main text, figures and Supplementary Information. S.S. and P.T. conceived, designed and performed the synchrotron experiments, processed and reconstructed the raw PPC-SRµCT scan data and wrote the PPC-SRµCT methods section. S.E.P. and J.L.M. performed, processed and reconstructed the µCT scan data. J.L.M. segmented all scan data and composed Supplementary Fig. 2. All authors provided a critical review of the manuscript and approved the final draft.

**Author Information** The synchrotron data will be made available through the European Synchrotron Radiation Facility (ESRF) palaeontology database (<http://paleo.esrf.eu>). To allow exploration of the scan data, reduced and compressed image stack movies have been deposited in the Dryad Digital Repository (<http://dx.doi.org/10.5061/dryad.0003s>). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.E.P. ([spierce@rvc.ac.uk](mailto:spierce@rvc.ac.uk)).

## METHODS

**PPC-SR $\mu$ CT scan protocol.** To maximize contrast of the structures, a 4–9-m propagation distance was used to make possible the use of propagation phase-contrast despite the large size (for example,  $15 \times 10 \times 5$  cm) of the samples and the high energy. As the samples were very dense owing to metallic oxide infilling, the specimens were scanned using the high energy white beam of the ID19 W150 wiggler at a gap of 38–80 mm filtered, with 3–9 mm of copper and 60 mm of aluminium, resulting in a beam with a relatively narrow bandwidth ( $83.7\text{--}133.2 \pm 15$  keV full-width at half-maximum). To homogenize absorption of the heterogeneous samples and to improve contrast, the fossil specimens were put in a 10-cm cylinder filled with aluminium/glass balls of 2 mm, and a beam profiler was used to compensate for absorption. The addition of the balls and the beam profiler provide the 60 mm of aluminium presented above for the beam filtering. This scanning protocol, derived from one originally developed for a high-energy monochromatic beam<sup>17</sup>, allows working with relatively low energy (when compared with the energy necessary to reach the usual 10% of transmission through the sample) while maximizing the sensitivity and the contrast (both absorption and phase) in the recorded data. Moreover, as the quantity of material crossed by the beam is nearly the same everywhere, this approach removes most of the beam hardening effect, except on the very dense metallic oxide particles.

A half-acquisition scanning geometry (off-axis centre of rotation) with 5,000 projections of 0.1 s over  $360^\circ$  was used to cover a lateral field of view of 100 mm with an isotropic voxel size of  $29.4 \mu\text{m}$ . In addition, a single crystal LuAG:Ce scintillator of  $750 \mu\text{m}$  thick was used, coupled through a lenses system with a CCD FReLoN (Fast Readout Low Noise) camera. To cover the full vertical field of view, series of scans were performed with vertical displacement between each scan in such a way that every position of the sample was scanned twice. This approach is used to reduce ring artefacts, as well as to minimize the artefacts due to differences in power and spectrum in the beam along its vertical profile. Following this, a single distance phase-retrieval algorithm<sup>31</sup> was used to maximize

the contrast of the structures in the data. Ring artefacts were corrected both on the radiographs before reconstruction by using filtered averages of all the projections, and after the reconstruction using a ring filter<sup>32</sup>. Data from all the scans were then combined into a single volume with moving average protocol to go from one scan to the next one. A three-dimensional unsharp mask was applied to the reconstructed data to retrieve the smallest details<sup>33</sup>. The scan data were scaled isotropically to  $58.8 \mu\text{m}$  using binning to reduce file size for segmentation while preserving a high level of resolution and the sensitivity of the original data set with an enhanced signal to noise ratio.

**$\mu$ CT scan protocol.** The specimen used in this study was scanned on an X-Tek HMX 225 keV system with a flat panel detector. The voxel size was set to  $90 \mu\text{m}$  using 175 kV and 127  $\mu\text{A}$ . Owing to the dense material, a 0.5 mm copper filter was used to filter out the low-energy X-rays and reduce beam-hardening effects. The scan data were corrected for ring artefacts and beam hardening using CT-Pro reconstruction software and the image stack was created using VGStudioMax.

**Segmentation and rendering.** All scan data were segmented in Mimics software (Materialize) and rendered in Autodesk 3D Studio Max. Fossil elements were linked and assigned object colours for visualization and movie generation.

**Specimen information.** *Ichthyostega* and *Acanthostega* specimens are housed in the Geological Museum, Copenhagen (MGUH), and *Pederpes* in the Hunterian Museum, Glasgow (GLAHMS).

31. Paganin, D., Mayo, S. C., Gureyev, T. E., Miller, P. R. & Wilkins, S. W. Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *J. Microsc.* **206**, 33–40 (2002).
32. Lyckegaard, A., Johnson, G. & Tafforeau, P. Correction of ring artifacts in X-ray tomographic images. *Int. J. Tomograph. Stat.* **18**, 1–9 (2011).
33. Sanchez, S., Ahlberg, P. E., Trinajstić, K., Mirone, A. & Tafforeau, P. Three dimensional synchrotron virtual paleohistology: a new insight into the world of fossil bone microstructures. *Microsc. Microanal.* **18**, 1095–1105 (2012).

# Biodiversity decreases disease through predictable changes in host community competence

Pieter T. J. Johnson<sup>1</sup>, Daniel L. Preston<sup>1</sup>, Jason T. Hoverman<sup>2</sup> & Katherine L. D. Richgels<sup>1</sup>

Accelerating rates of species extinctions and disease emergence underscore the importance of understanding how changes in biodiversity affect disease outcomes<sup>1–3</sup>. Over the past decade, a growing number of studies have reported negative correlations between host biodiversity and disease risk<sup>4–8</sup>, prompting suggestions that biodiversity conservation could promote human and wildlife health<sup>9,10</sup>. Yet the generality of the diversity–disease linkage remains conjectural<sup>11–13</sup>, in part because empirical evidence of a relationship between host competence (the ability to maintain and transmit infections) and the order in which communities assemble has proven elusive. Here we integrate high-resolution field data with multi-scale experiments to show that host diversity inhibits transmission of the virulent pathogen *Ribeiroia ondatrae* and reduces amphibian disease as a result of consistent linkages among species richness, host composition and community competence. Surveys of 345 wetlands indicated that community composition changed nonrandomly with species richness, such that highly competent hosts dominated in species-poor assemblages whereas more resistant species became progressively more common in diverse assemblages. As a result, amphibian species richness strongly moderated pathogen transmission and disease pathology among 24,215 examined hosts, with a 78.4% decline in realized transmission in richer assemblages. Laboratory and mesocosm manipulations revealed an approximately 50% decrease in pathogen transmission and host pathology across a realistic diversity gradient while controlling for host density, helping to establish mechanisms underlying the diversity–disease relationship and their consequences for host fitness. By revealing a consistent link between species richness and community competence, these findings highlight the influence of biodiversity on infection risk and emphasize the benefit of a community-based approach to understanding infectious diseases.

Worldwide, ecological systems continue to undergo dramatic changes in biodiversity that affect a range of community and ecosystem processes<sup>1,3</sup>. Recently, biodiversity changes have also been linked to shifts in disease risk for humans and wildlife<sup>9,10,14</sup>. Because many pathogens infect multiple host species that vary in their competence<sup>2,15</sup>, host community composition can acutely influence disease outcomes. If more diverse assemblages support a greater fraction of low-competency hosts, biodiversity losses have the potential to increase disease risk ('dilution effect')<sup>16</sup>. Although support for this hypothesis has accumulated for a growing list of human, plant and wildlife diseases<sup>4,5,7,8,17–19</sup>, uncertainty persists over the generality of such patterns and their underlying mechanisms<sup>11–13</sup>.

The outcome of biodiversity changes for infectious disease risk ultimately depends on the specific order in which species are added to or lost from a community relative to their competence for supporting infection<sup>20</sup>. If increases in diversity lead predictably to the addition of low-competency species—thereby lowering the community's capacity to support infection ('community competence')—a dilution effect will occur. In contrast, the addition of high competency hosts with increasing richness can cause an amplification effect<sup>16,20</sup>. Although ecologists have long-recognized that communities are nonrandomly

structured<sup>21</sup>, the influence of naturally occurring patterns of community assembly and disassembly has only recently been more widely incorporated into studies of biodiversity loss<sup>22</sup>. Aside from simulation-based models<sup>20</sup>, however, this concept has received surprisingly little attention in disease research, and how community host competence and species richness co-vary in natural systems remains largely unexplored<sup>7</sup>.

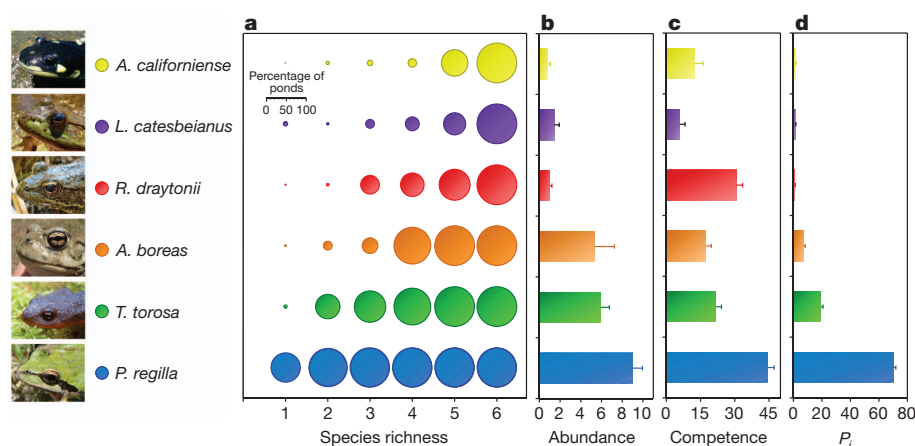
Here we combined high-resolution sampling of replicate host assemblages with multi-scale experiments to test the effects of host diversity on pathogen transmission and evaluate the role of functional changes in community competence in driving this pattern. We focused on interactions between pond-breeding amphibians and the multi-host trematode *Ribeiroia ondatrae*, which causes mortality and severe malformations<sup>23</sup>. Pond systems are well suited to address questions involving assembly because they provide replicate assemblages and support an experimentally tractable gradient of host richness (here, 1 to 6 species). Over three years, we sampled amphibian richness and composition within 345 wetlands across a 758,100 hectare region in California, USA (Supplementary Fig. 1) and assessed pathology (malformations) in 24,215 hosts. Experimentally derived estimates of host competence were combined with information on species composition to test whether increases in host richness lead predictably to decreases in community competence. Because predictions of the dilution effect often involve changes in transmission<sup>16</sup>, or the dynamic ability of pathogens to move between hosts, simple comparisons of host richness and infection prevalence or intensity may fail to adequately test this framework. We therefore derived empirical estimates of realized transmission (see Supplementary Methods) by quantitatively linking information on infection from 17,516 snails, which function as intermediate hosts, to *R. ondatrae* load within 4,520 amphibians. We complemented field surveys with experiments that directly tested for diversity-driven changes in pathogen transmission within realistic host assemblages.

Among naturally occurring assemblages, amphibian species composition changed consistently along a gradient of species richness, leading to a nested pattern in which low-diversity communities formed near-perfect subsets of more diverse assemblages (matrix temperature = 20.47°;  $P < 0.001$ ; Fig. 1a). Estimates of host competence correlated positively with field-based measures of both a species' occurrence and its abundance when present, leading to variation in each species' contribution to 'community competence' (see Methods and Fig. 1). Accordingly, the most-competent amphibian host (*Pseudacris regilla*) was also the most common species with progressive decreases in the fraction of highly competent hosts in more diverse assemblages (Fig. 2a). Integrating host species' competence with their relative abundance, we estimated a 35.7% decrease in community competence over the observed richness gradient (generalized linear model (GLM)  $F_{1, 286} = 7.72$ ,  $P = 0.006$ , Fig. 2b).

The nested structure of host assemblages led to strong differences in realized transmission and disease pathology with changes in host richness. Among host populations, mean infection load explained ~75% of variation in malformation frequency (GLM,  $t = 24.46$ ,  $P < 0.0001$ ;

<sup>1</sup>Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309, USA. <sup>2</sup>Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana 47907, USA.





**Figure 1 | Consistent linkages among species richness, community composition and the traits of individual amphibian host species across 345 sampled wetlands.** **a**, The percentage of wetlands supporting each species is represented by the size of the circle within each richness level (nested structure; observed matrix temperature = 20.47°; average null model matrix temperature = 69.13°;  $P < 0.001$ ). The host species were *Ambystoma californiense*, *Lithobates catesbeianus* (also known as *Rana catesbeiana*), *Rana draytonii*, *Anaxyrus boreas*, *Taricha torosa* and *Pseudacris regilla*. **b–d**, Shown for each host species is the mean abundance (number  $m^{-2}$ ) when present

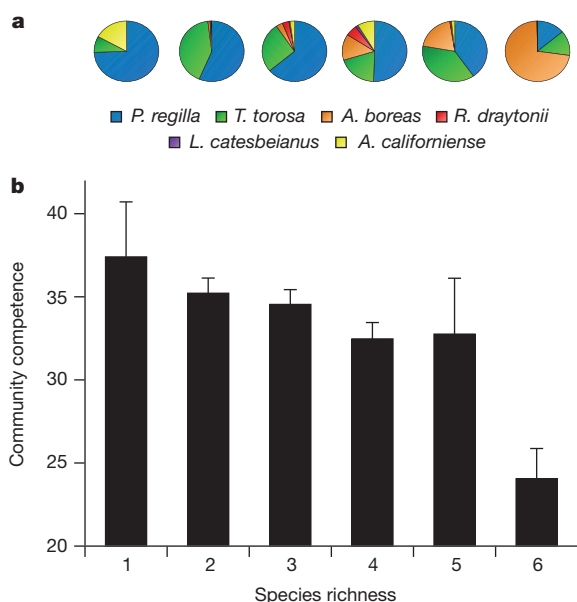
(**b**), its laboratory-measured competence for supporting *R. ondatrae* (**c**), and an index of each host species' contribution to community competence ( $P_i$ , which combines the fraction of wetlands occupied by a host, its relative abundance when present, and host competence—scaled between 0 and 100%) (**d**). All error bars represent standard error (s.e.). Species occurrence, abundance, competency and body size (not shown) all loaded strongly ( $> |0.87|$ ) on a single principal component (eigenvalue = 3.5, 89% of variation). Images were provided by G. Nafis (*A. californiense*) and D. Preston (all others).

Fig. 3a), which affected up to 90% of hosts in some populations (Supplementary Discussion). However, amphibian species richness strongly moderated transmission between snails and amphibians ( $R^2 = 0.48$ ,  $F_{3, 132} = 40.33$ , richness  $\times$  infected snail density  $P < 0.0001$ ), with a 78.4% reduction in realized transmission and a 52.6% decrease in pathology in more diverse assemblages (Fig. 3b and Supplementary Discussion). Thus, infected snail density positively predicted amphibian infection, but infection success between snails and amphibians

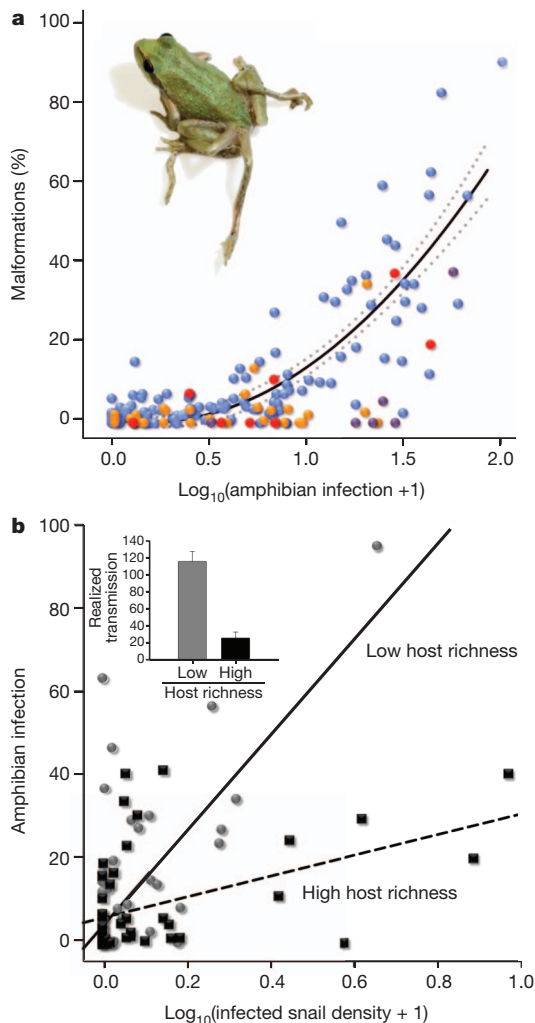
decreased with progressive increases in amphibian richness (Fig. 3b). The best-supported models included main effects and interactions between infected snail density and either amphibian richness ( $\Delta$ Akaike's information criterion (AIC) = 0,  $w_i = 0.76$ ) or the occurrence of each host species ( $\Delta$ AIC = 4.08,  $w_i = 0.10$ ) (Supplementary Discussion). Results were robust to whether we analysed data using individual hosts nested within wetlands or averaged among hosts from each site (Supplementary Discussion). Among wetlands for which we obtained infection data from all amphibian host species, community competence and its interaction with infected snail density explained 89% of the variance in total *R. ondatrae* load (summed among all host species) ( $R^2 = 0.89$ ,  $F_{3, 21} = 57.93$ ,  $P < 0.0001$ ). This model was a better fit to the data than those with amphibian richness, host density, or the occurrence of particular host species ( $\Delta$ AIC = 10.7).

Experimental manipulations reinforced field observations and provided insight into the relative effects of host density and composition on transmission. In laboratory manipulations, increases in host richness from one to three species caused a 64% reduction in transmission (GLM, richness:  $-9.648 \pm 2.265$ ,  $P < 0.0001$ ; Fig. 4a). This effect was not entirely attributable to changes in the density of the most competent host, as infection decreased by 28% even when *P. regilla* density was fixed and other species were introduced additively. In outdoor mesocosms designed to mimic natural assemblages, both total infection (summed among host individuals and species) and per capita infections in *P. regilla* decreased by approximately 50% between the lowest (one species) and highest richness (four species) treatments (total: GLM  $F_{2, 22} = 18.19$ ,  $P < 0.0001$ ; *P. regilla*: Generalized linear mixed model (GLMM)  $t = -3.6761$ ,  $P = 0.0013$ ; Fig. 4b and Supplementary Discussion). Correspondingly, increases in richness caused a decline in disease pathology (Fig. 4c).

By integrating information on community structure, host competence and pathogen infection from a large number of replicate assemblages, this study provides empirical evidence of a functional link between biodiversity and pathogen transmission. Our findings help to strengthen the conceptual foundation underlying the dilution effect by demonstrating that strongly nested assemblages can lead to a predictable relationship between host richness and community competence, in which low diversity assemblages support a greater proportion of highly competent host species<sup>6</sup>. Combined results of experiments and field-based measurements of realized transmission

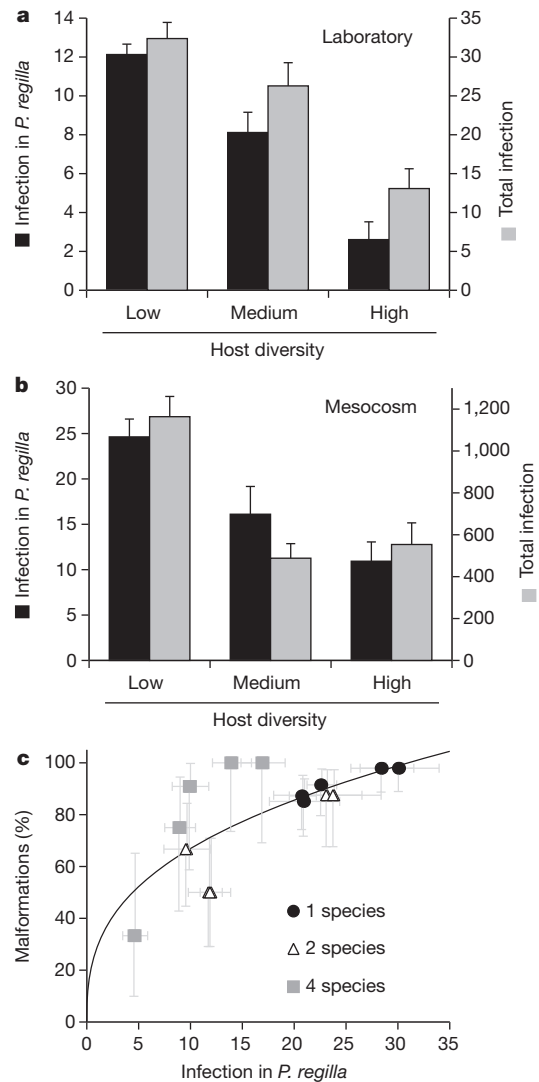


**Figure 2 | Influence of amphibian species richness on the capacity of communities to support parasite infection in naturally occurring wetlands.** **a**, **b**, Relative abundance of each host species (**a**) and mean community competence (**b**) ( $\pm 1$  s.e.) as a function of amphibian species richness. Community competence integrates information on each host species' competence and relative abundance within the community, providing a metric of the potential for communities at each richness level to support infection (see Supplementary Methods). This measure scales between the lowest (*L. catesbeianus* = 1) and highest (*P. regilla* = 44.34) observed competence values ( $c_{\min}$  and  $c_{\max}$ ).



**Figure 3 | Effects of parasite infection on amphibian malformations and of host richness on parasite transmission.** **a**, Relationship between *R. ondatrae* infection and malformation frequency in anuran host populations. Malformation data were assessed through inspections of 24,215 amphibians, whereas infection data were derived from 4,520 necropsied hosts (see Fig. 1 for host colour scheme). Dotted lines represent 95% confidence (GLM with quasibinomial distribution,  $\log_{10} (R. ondatrae + 1)$ ,  $t = 24.46$ ,  $P < 0.0001$ ; pseudo- $R^2 = 74.74$ ,  $n = 291$  species  $\times$  wetland combinations). **b**, Effects of amphibian species richness on realized *R. ondatrae* transmission between snail intermediate hosts and amphibians. The y axis displays *R. ondatrae* load (mean per *P. regilla*) and the x-axis displays the  $\log_{10}$  (density of infected snails + 1). Amphibian richness interacted significantly with infected snail density to determine *R. ondatrae* load, with decreases in the transmission coefficient between low richness (1–3 species; grey circles and solid line) and high richness (4–6 species; black squares and dashed line) ( $R^2 = 0.48$ ,  $F_{3, 132} = 40.3314$ ,  $P < 0.0001$ ,  $n = 136$  site visits; 1,798 necropsied amphibians and 11,041 dissected snails). Inset depicts the realized transmission coefficients ( $\pm 1$  s.e.) in low and high richness assemblages. Image provided by D. Herasimtschuk.

indicate that this relationship leads directly to a reduction in infection success at higher host richness, even after controlling for changes in host density. Nonrandom structuring of ecological communities has proven similarly important for explaining the relationship between biodiversity and ecosystem functioning<sup>22</sup>, but has rarely been demonstrated for studies of infectious disease in a community context<sup>20,24</sup>. One possible explanation for the negative relationship between a host's competence and its assembly order is that defences are costly and may incur trade-offs with resource investment in reproduction or dispersal<sup>25</sup>. Indeed, studies in both eco-immunology and conservation support linkages between a species life history traits (for example, 'pace of life') and its vulnerability to infection or extinction, respectively<sup>26,27</sup>.



**Figure 4 | Experimental effects of host diversity on parasite transmission and host pathology.** **a**, **b**, Effects of host richness on per capita *R. ondatrae* infection in *P. regilla* (black bars; mean  $\pm 1$  s.e.) and total infection across all host species (grey bars; mean  $\pm 1$  s.e.) in laboratory (**a**) and mesocosm (**b**) experiments. Low, medium and high richness represent one, two and three species, respectively, in the laboratory experiment, and one, two and four species, respectively, in the mesocosm experiment. **c**, Relationship between average *R. ondatrae* infection ( $\pm 1$  s.e.) and malformation frequency ( $\pm 95\%$  confidence interval) in *P. regilla* among mesocosm host richness treatments (GLMM with binomial distribution, effect of richness  $Z = -2.71$ ,  $P = 0.029$ ).

These findings also advocate care in specifying what epidemiological outcomes are associated with diversity changes. Here, host diversity interacted strongly with infected snail density to determine amphibian infection, indicating that the richness of hosts moderated the realized transmission success (that is,  $\beta'$ ) of parasites in moving between snails and amphibians, rather than through alternative mechanisms such as controlling the abundance of available hosts<sup>16</sup>. For many multi-host and vector-borne infections, the density of vectors, intermediate stages and susceptible hosts will vary spatially and temporally due to a range of environmental factors (for example, climate, resources or habitat), independent of local diversity. This emphasizes the importance of field-based estimates of infection pressure and suggests that studies relying on static comparisons of host richness and infection prevalence or intensity may fail to capture the dynamic effects of diversity on transmission. Finally, it is important to note that the effects of diversity will also depend on the specific response metric under consideration. For instance, host diversity and parasite diversity will often correlate

positively due to enhanced colonization opportunities<sup>28</sup>, but parasite richness is not equivalent to disease risk, particularly when parasites vary in virulence and relative abundance<sup>29</sup>.

Our results lend mechanistic insight into how host diversity can reduce disease risk through predictable co-variation between species-level traits and community assembly. In light of mounting evidence that higher biodiversity can buffer against pathogen exposure in human, wildlife and plant disease systems<sup>4,6,8,10,17–19</sup>, preserving functional diversity—including both genetic diversity and community richness—has the potential to ameliorate pathogen transmission and offer a novel, cost-effective approach to disease management.

## METHODS SUMMARY

**Field surveys.** Between 2009 and 2011, we estimated the abundance and species composition of amphibians among 345 wetlands in California, USA (Supplementary Fig. 1). At sites with *Planorbella trivolvis* (also known as *Helisoma trivolvis*), the requisite snail host for *R. ondatrae*, we conducted a second visit to quantify malformation frequency and infection load (Supplementary Methods, 678 total site visits). *R. ondatrae*-infected snail density was estimated as the product of snail density and infection prevalence. We used bootstrapping and rarefaction analyses to validate parasite abundance and amphibian richness estimates (Supplementary Figs 3 and 4) and calculated nestedness using the nestedness temperature calculator (Supplementary Methods).

**Experimental studies.** In 2 litre containers, we tested how the number of amphibian hosts and host species richness affected *R. ondatrae* infection success. To increase ecological realism, we conducted a 2 × 3 factorial experiment manipulating infection and amphibian host diversity within outdoor mesocosms. Both sets of experiments compared additive and substitutive designs (Supplementary Methods).

**Laboratory measures of host competence.** We measured host competence by exposing individuals of each host species to *R. ondatrae* cercariae during early limb development and quantifying the proportion of encysted parasites remaining immediately following metamorphosis (see Supplementary Methods and ref. 23).

**Full Methods** and any associated references are available in the online version of the paper.

Received 26 August; accepted 31 December 2012.

- Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
- Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).
- Naeem, S., Duffy, J. E. & Zavaleta, E. The functions of biological diversity in an age of extinction. *Science* **336**, 1401–1406 (2012).
- Clay, C. A., Lehmer, E. M., Jeor, S. S. & Dearing, M. D. Sin Nombre virus and rodent species diversity: a test of the dilution and amplification hypotheses. *PLoS ONE* **4**, e6467 (2009).
- Haas, S. E., Hooten, M. B., Rizzo, D. M. & Meentemeyer, R. K. Forest species diversity reduces disease risk in a generalist plant pathogen invasion. *Ecol. Lett.* **14**, 1108–1116 (2011).
- LoGiudice, K., Ostfeld, R. S., Schmidt, K. A. & Keesing, F. The ecology of infectious disease: Effects of host diversity and community composition on Lyme disease risk. *Proc. Natl Acad. Sci. USA* **100**, 567–571 (2003).
- Allan, B. F. *et al.* Ecological correlates of risk and incidence of West Nile virus in the United States. *Oecologia* **158**, 699–708 (2009).
- Ezenwa, V. O., Godsey, M. S., King, R. J. & Gupta, S. C. Avian diversity and West Nile virus: testing associations between biodiversity and infectious disease risk. *Proc. R. Soc. B* **273**, 109–117 (2006).
- Keesing, F. *et al.* Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* **468**, 647–652 (2010).

- Ostfeld, R. S. & Keesing, F. Effects of host diversity on infectious disease. *Annu. Rev. Ecol. Syst.* **43**, 157–182 (2012).
- Randolph, S. E. & Dobson, A. D. M. Pangloss revisited: a critique of the dilution effect and the biodiversity-buffers-disease paradigm. *Parasitology* **139**, 847–863 (2012).
- Hamer, G. L. *et al.* Fine-scale variation in vector host use and force of infection drive localized patterns of West Nile virus transmission. *PLoS ONE* **6**, e23767 (2011).
- Wood, C. L. & Lafferty, K. D. Biodiversity and disease: a synthesis of ecological perspectives on Lyme disease transmission. *Trends Ecol. Evol.* <http://dx.doi.org/10.1016/j.tree.2012.10.011> (23 November 2012).
- Raymundo, L. J., Halford, A. R., Maypa, A. P. & Kerr, A. M. Functionally diverse reef-fish communities ameliorate coral disease. *Proc. Natl Acad. Sci. USA* **106**, 17067–17070 (2009).
- Kilpatrick, A. M. Globalization, land use, and the invasion of West Nile virus. *Science* **334**, 323–327 (2011).
- Keesing, F., Holt, R. D. & Ostfeld, R. S. Effects of species diversity on disease risk. *Ecol. Lett.* **9**, 485–498 (2006).
- Suzán, G. *et al.* Experimental evidence for reduced rodent diversity causing increased Hantavirus prevalence. *PLoS ONE* **4**, e5461 (2009).
- Searle, C. L., Biga, L. M., Spatafora, J. W. & Blaustein, A. R. A dilution effect in the emerging amphibian pathogen *Batrachochytrium dendrobatidis*. *Proc. Natl Acad. Sci. USA* **108**, 16322–16326 (2011).
- Johnson, P. T. J. *et al.* Species diversity reduces parasite infection through cross-generational effects on host abundance. *Ecology* **93**, 56–64 (2012).
- Ostfeld, R. S. & LoGiudice, K. Community disassembly, biodiversity loss, and the erosion of an ecosystem service. *Ecology* **84**, 1421–1427 (2003).
- Ricklefs, R. E. Community diversity: relative roles of local and regional processes. *Science* **235**, 167–171 (1987).
- Bracken, M. E. S., Friberg, S. E., Gonzalez-Dorantes, C. A. & Williams, S. L. Functional consequences of realistic biodiversity changes in a marine ecosystem. *Proc. Natl Acad. Sci. USA* **105**, 924–928 (2008).
- Johnson, P. T. J. *et al.* Living fast and dying of infection: host life history drives interspecific variation in infection and disease risk. *Ecol. Lett.* **15**, 235–242 (2012).
- Graham, S. P., Hassan, H. K., Burkett-Cadena, N. D., Guyer, C. & Unnasch, T. R. Nestedness of ectoparasite-vertebrate host networks. *PLoS ONE* **4**, e7873 (2009).
- Ricklefs, R. E. & Wikelski, M. The physiology/life-history nexus. *Trends Ecol. Evol.* **17**, 462–468 (2002).
- Lee, K. A., Wikelski, M., Robinson, W. D., Robinson, T. R. & Klasing, K. C. Constitutive immune defences correlate with life-history variables in tropical birds. *J. Anim. Ecol.* **77**, 356–363 (2008).
- Purvis, A., Gittleman, J. L., Cowlshaw, G. & Mace, G. M. Predicting extinction risk in declining species. *Proc. R. Soc. Lond. B* **267**, 1947–1952 (2000).
- Hechinger, R. F. & Lafferty, K. D. Host diversity begets parasite diversity: bird final hosts and trematodes in snail intermediate hosts. *Proc. R. Soc. B* **272**, 1059–1066 (2005).
- Johnson, P. T. J. & Hoverman, J. T. Parasite diversity and coinfection determine pathogen infection success and host fitness. *Proc. Natl Acad. Sci. USA* **109**, 9006–9011 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank B. Hoyer, J. Koprivnikar, K. Medley, J. Rohr and especially Y. Springer for editorial suggestions; S. Johnson for valuable statistical advice; M. Baragona, I. Buller, K. Gietzen, B. Goodman, J. Jenkins, E. Kellermanns, B. LaFonte, T. McDevitt-Galles, J. McFarland and S. Paull for assistance in collecting data; and East Bay Regional Parks, East Bay Municipal Utility District, Santa Clara County Parks, Hopland Research and Extension Center, Blue Oak Ranch Reserve, California State Parks, The Nature Conservancy, Open Space Authority and Mid-peninsula Open Space for access to properties and logistical support. This work was supported through funds from the US National Science Foundation (DEB-0841758, DEB-1149308), the National Geographic Society, and the David and Lucile Packard Foundation.

**Author Contributions** P.T.J.J. designed the study, D.L.P., K.L.D.R. and P.T.J.J. collected the data, P.T.J.J. and J.T.H. analysed the data, and all authors wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.T.J.J. ([pieter.johnson@colorado.edu](mailto:pieter.johnson@colorado.edu)).



## METHODS

**Field surveys and parasite assessment.** Over 3 years, we used standardized methods to survey amphibian abundance and species composition in 345 wetlands from across a 758,100 hectare region of California, USA (Supplementary Fig. 1 and Supplementary Methods). At sites with *P. trivolvus*, the requisite snail host for *R. ondatrae*, we conducted a second late-summer visit to quantify malformations and collect amphibians for necropsy (678 total site visits). *R. ondatrae*-infected snail density was estimated by multiplying snail density by infection prevalence (Supplementary Methods). To quantify *R. ondatrae* infection in amphibians, we necropsied 10–15 recently metamorphosed *P. regilla* per site<sup>30,31</sup>. We focused on *P. regilla* because this amphibian host is widespread, locally abundant and sensitive to parasite-induced malformations, thereby offering a consistent basis for cross-site comparisons. At a subset of sites ( $n = 25$ ), however, we necropsied all co-occurring amphibian species and used these data to generate estimates of total *R. ondatrae* abundance, or the product of each host's relative abundance and its average *R. ondatrae* infection intensity. We used bootstrapping and rarefaction analyses to validate the efficacy of parasite abundance and amphibian richness estimates (Supplementary Figs 2 and 3).

**Estimates of host competence.** We estimated host competence, or the ability of *R. ondatrae* to infect and persist within each host species, of the six amphibian species present at our field sites by exposing larvae to *R. ondatrae* cercariae in 1.5 litre containers<sup>32</sup>. Cercariae were obtained from wild-caught *P. trivolvus* and administered to amphibians within 4 h of release. Host species were collected as eggs and allowed to develop to stage 28 (anurans<sup>33</sup>) or stage 2T (newts<sup>34</sup>) before exposure. For the two endangered species under federal protection, we used closely related congeners as surrogates (*Rana aurora* and *Ambystoma tigrinum*)<sup>35,36</sup>. Hosts were necropsied at metamorphosis to quantify metacercariae, and we calculated competence as the percentage of parasites detected relative to the number administered<sup>16</sup>. For field analyses, we built upon this approach by calculating 'community competence' ( $p$ ) for the amphibian assemblage at each wetland (modified from<sup>37</sup>):  $p = \frac{\sum_{i=1}^n c_i S_i}{\sum_{i=1}^n S_i}$  where  $c_i$  is the competence of species  $i$  and  $S_i$  is its abundance (which becomes relative abundance when divided by the total number of amphibian larvae captured).

**Laboratory and mesocosm experiments.** In laboratory experiments (2 litres), we tested how the number of hosts (1, 2 or 3 *P. regilla*) and host community composition (1, 2 or 3 species) affected both the infection per host and the total number of successful parasites. For the diversity portion, each replicate contained three individuals representing one species (three treatments involving three individuals of *P. regilla*, *Rana cascadae*, or *L. catesbeianus*), two species (two treatments involving two *P. regilla* and either one *R. cascadae* or one *L. catesbeianus*) and three species (one individual of each species) ( $n = 15$  replicates per treatment). Over 5 days, we added 60 *R. ondatrae* cercariae to each container; hosts were euthanized, measured and necropsied after 48 h. To increase ecological realism, we conducted a  $2 \times 3$  factorial experiment manipulating *R. ondatrae* infection (yes or no) and host diversity (1, 2 or 4 host species) within outdoor mesocosms (Supplementary Methods). Diversity treatments contained one species (60 *P. regilla*), two species (30 *P. regilla* and 30 *A. boreas*), or four species (15 *P. regilla*, 15 *A. boreas*, 15 *T. torosa* and 6 *L. catesbeianus*) (substitutive design). The design was informed by field data such that the four most common species (85% of observations) were included in the high diversity treatment, and the most ubiquitous species (*P. regilla*) was the focal host in the one species treatment. In the density experiment, mesocosms contained 15, 30 or 60 *P. regilla*. All treatments were replicated five times. Over 8 days, we added 5,324 cercariae ( $\pm 162$  s.e.) to each mesocosm (Supplementary Methods) and collected amphibian hosts as they reached stage 42 (*P. regilla* and *A. boreas*) or after 40 days (*L. catesbeianus* and *T. torosa*).

**Field analyses.** We used the program Aninhado<sup>38</sup> to test for nestedness using the nested temperature calculator<sup>39</sup> and compared the observed metric score to 1,000 permutations of a null model. To evaluate the link between *R. ondatrae* infection

and malformations, we used a generalized linear model with a quasibinomial response and included a fixed effect for species identity<sup>40,41</sup> (Supplementary Methods). Among sites that supported *R. ondatrae*, we tested how host species richness affected parasite transmission by incorporating information on infected snail density ( $\log_{10}$ -transformed). We expected the density of infected snails to positively predict amphibian infection (realized transmission) and sought to identify how host diversity moderated this relationship. This analysis was performed both at the site level, in which infection in frogs was averaged, and at the individual host level, in which frogs were nested within site using a generalized linear mixed model with a negative binomial distribution. Alongside infected snail density, which was included in all models as an a priori measure of parasite infection pressure, we included measurements of host diversity (richness, Shannon diversity and evenness)<sup>42</sup>, the presence of each host species, and the density of amphibian larvae ( $\log_{10}$ -transformed). Because we expected host species richness or composition to function as moderators of transmission success, we included an interaction term between infected snail density and each predictor. We assessed among models using Akaike's information criterion and model weights<sup>43</sup>. For wetlands with necropsy data from all amphibian species ( $n = 25$ ), we repeated this procedure with community competence and its interaction with infected snail density.

**Laboratory and mesocosm analyses.** To test how host density and species richness affected per capita infections, we used generalized linear mixed models in which hosts were nested within replicates and parasite counts were modelled using a negative binomial distribution<sup>40,41</sup>. This approach was also used to test how assemblage structure affected malformations and host development. To test effects on total amphibian infection (summed among host species), we used generalized linear models with Gaussian or Poisson distributions. By comparing the diversity and density treatments, we also contrasted the effects of diversity between additive and substitutive designs (we compared the effects of changes in species composition both when the total number of hosts was held constant (substitutive) and when the abundance of the focal host was held constant and other species were added sequentially (additive)).

30. Johnson, P. T. J. & Buller, I. D. Parasite competition hidden by correlated coinfection: using surveys and experiments to understand parasite interactions. *Ecology* **92**, 535–541 (2011).
31. Johnson, P. T. J. & Hartson, R. B. All hosts are not equal: explaining differential patterns of malformations in an amphibian community. *J. Anim. Ecol.* **78**, 191–201 (2009).
32. Johnson, P. T. J. *et al.* Living fast and dying of infection: host life history drives interspecific variation in infection and disease risk. *Ecol. Lett.* **15**, 235–242 (2012).
33. Gosner, K. L. A simplified table for staging anuran embryos and larvae with notes and identification. *Herpetologica* **16**, 183–190 (1960).
34. Wong, C. J. & Liversage, R. A. Limb developmental stages of the newt *Notophthalmus viridescens*. *Int. J. Dev. Biol.* **49**, 375–389 (2005).
35. Hillis, D. M. & Wilcox, T. P. Phylogeny of the New World true frogs (*Rana*). *Mol. Phylogenet. Evol.* **34**, 299–314 (2005).
36. Shaffer, H. B., Clark, J. M. & Kraus, F. When molecules and morphology clash: a phylogenetic analysis of the North American ambystomatid salamanders (Caudata: Ambystomatidae). *Syst. Zool.* **40**, 284–303 (1991).
37. Mitchell, C. E., Reich, P. B., Tilman, D. & Groth, J. V. Effects of elevated CO<sub>2</sub>, nitrogen deposition, and decreased species diversity on foliar fungal plant disease. *Glob. Change Biol.* **9**, 438–451 (2003).
38. Guimarães, P. R. Jr & Guimarães, P. Improving the analyses of nestedness for large sets of matrices. *Environ. Modell. Softw.* **21**, 1512–1513 (2006).
39. Atmar, W. & Patterson, B. D. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* **96**, 373–382 (1993).
40. Bolker, B. M. *et al.* Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* **24**, 127–135 (2009).
41. Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A. & Smith, G. M. *Mixed Effects Models and Extensions in Ecology with R* (Springer, 2009).
42. Magurran, A. E. *Measuring Biological Diversity* (Blackwell Publishing, 2004).
43. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference* (Springer, 2002).

# Finding the sources of missing heritability in a yeast cross

Joshua S. Bloom<sup>1,2</sup>, Ian M. Ehrenreich<sup>1,3</sup>, Wesley T. Loo<sup>1,2</sup>, Thúy-Lan Võ Lite<sup>1,2</sup> & Leonid Kruglyak<sup>1,4,5</sup>

**For many traits, including susceptibility to common diseases in humans, causal loci uncovered by genetic-mapping studies explain only a minority of the heritable contribution to trait variation. Multiple explanations for this ‘missing heritability’ have been proposed<sup>1</sup>. Here we use a large cross between two yeast strains to accurately estimate different sources of heritable variation for 46 quantitative traits, and to detect underlying loci with high statistical power. We find that the detected loci explain nearly the entire additive contribution to heritable variation for the traits studied. We also show that the contribution to heritability of gene–gene interactions varies among traits, from near zero to approximately 50 per cent. Detected two-locus interactions explain only a minority of this contribution. These results substantially advance our understanding of the missing heritability problem and have important implications for future studies of complex and quantitative traits.**

Individuals within species show heritable variation for many traits of biological and medical interest. Most heritable traits follow complex inheritance patterns, with multiple underlying genetic factors<sup>2,3</sup>. Finding these factors has been a central focus of modern genetic research in humans, as well as in model organisms and agriculturally important species<sup>4–7</sup>. Recent work, most notably genome-wide association studies (GWAS) in humans, has underscored the problem of missing heritability: although many genetic loci have been identified for a wide range of traits, these typically explain only a minority of the heritability of each trait, implying the existence of other, undiscovered genetic factors<sup>1</sup>.

Multiple non-mutually exclusive explanations have been proposed for missing heritability<sup>1</sup>. One possibility is that the undiscovered factors could have effects that are too small to be detected with current sample sizes, or even too small to ever be individually detected with statistical significance<sup>8,9</sup>. The existence of many small-effect variants is supported by studies showing that a large proportion of heritable trait variation is tagged when all GWAS markers are considered simultaneously<sup>10,11</sup>. Because GWAS can only detect variants that are common in the population, another possibility is that the undiscovered variants are too rare in the population to be captured by GWAS<sup>12</sup>. One recent proposal<sup>13</sup> highlights the fact that non-additive interactions among loci (sometimes termed epistasis) may inflate heritability measures<sup>14</sup>. Other proposed contributions to missing heritability include structural variation, gene–environment interactions, parent-of-origin effects, heritable epigenetic factors and ‘entirely unforeseen sources’<sup>11,15</sup>. A better understanding of the sources of missing heritability is crucial for designing studies to find the missing components.

We set out to investigate these questions in the yeast *Saccharomyces cerevisiae*. We and others have previously used a cross between a laboratory strain and a wine strain to investigate the genetic basis of many complex traits, including global gene expression, protein abundance, telomere length, cell shape, gene-expression noise and drug sensitivity, and we have demonstrated missing heritability in this

system<sup>16,17</sup>. Recently, we used extreme quantitative trait locus mapping (X-QTL), a bulk segregant approach that uses pools of millions of cross progeny (segregants), to detect many loci underlying heritable trait variation<sup>18</sup>. We showed that for one trait, loci detected by X-QTL explained most of the heritability. However, pooled approaches do not allow direct estimates of heritability, the contribution of gene–gene interactions or locus effect sizes. Here we use a large panel of individually genotyped and phenotyped yeast segregants to accurately measure the heritable components of many quantitative traits, discover the underlying loci and examine the sources of missing heritability.

To estimate heritability and detect the underlying loci with high statistical power, we constructed a panel of 1,008 prototrophic haploid segregants from a cross between a laboratory strain and a wine strain (Supplementary Fig. 1 and Methods). These strains differ by 0.5% at the sequence level<sup>19</sup>. We sequenced the parent strains to high coverage and compared the sequences to define 30,594 high-confidence single-nucleotide polymorphisms that distinguish the strains and densely cover the genome. We obtained comprehensive individual genotype information for each of the 1,008 segregants by highly multiplexed short-read sequencing (Supplementary Fig. 1d).

We sought to accurately measure a large number of quantitative traits in the segregant panel. To do so, we implemented a high-throughput end-point colony size assay and measured growth in multiple conditions, including different temperatures, pHs and carbon sources, as well as addition of metal ions and small molecules<sup>16,18</sup> (Supplementary Fig. 1c). We defined each trait as end-point colony size normalized relative to growth on control medium (Methods), and obtained reproducible measurements with a strong heritable component for 46 traits (Supplementary Table 1). Most of the traits were only weakly correlated with each other (Supplementary Fig. 2).

Phenotypic variation in the segregant panel can be partitioned into the contribution of heritable genetic factors (broad-sense heritability) and measurement errors or other random environmental effects. Broad-sense heritability can, in turn, be partitioned into the contribution of additive genetic factors (narrow-sense heritability), dominance effects, gene–gene interactions and gene–environment interactions<sup>14</sup>. In our experiment, dominance effects are absent because the segregants are haploid, and gene–environment interactions for a given trait should also be absent as all the segregants are grown simultaneously under uniform conditions. Thus our estimates of broad-sense heritability include additive and gene–gene interaction components, whereas our estimates of narrow-sense heritability include only the additive component. The difference between the two heritability measures therefore provides an estimate of the contribution of gene–gene interactions.

We estimated broad-sense heritability from repeatability of trait measurements (Methods). Estimating narrow-sense heritability usually involves measuring phenotypic similarity for different degrees of relatedness. We took advantage of a recently developed genomic approach in which narrow-sense heritability is estimated by comparing phenotypic similarity among individuals with their actual genetic relatedness,

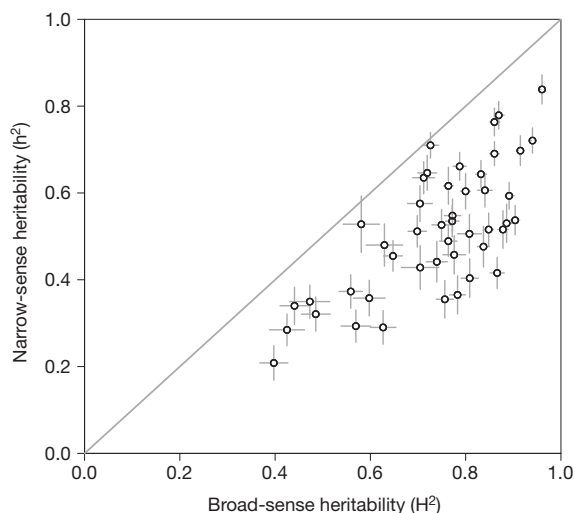
<sup>1</sup>Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08540, USA. <sup>2</sup>Department of Molecular Biology, Princeton University, Princeton, New Jersey 08540, USA.

<sup>3</sup>Molecular and Computational Biology Section, University of Southern California, Los Angeles, California 90089, USA. <sup>4</sup>Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08540, USA. <sup>5</sup>Howard Hughes Medical Institute, Princeton University, Princeton, New Jersey 08540, USA.

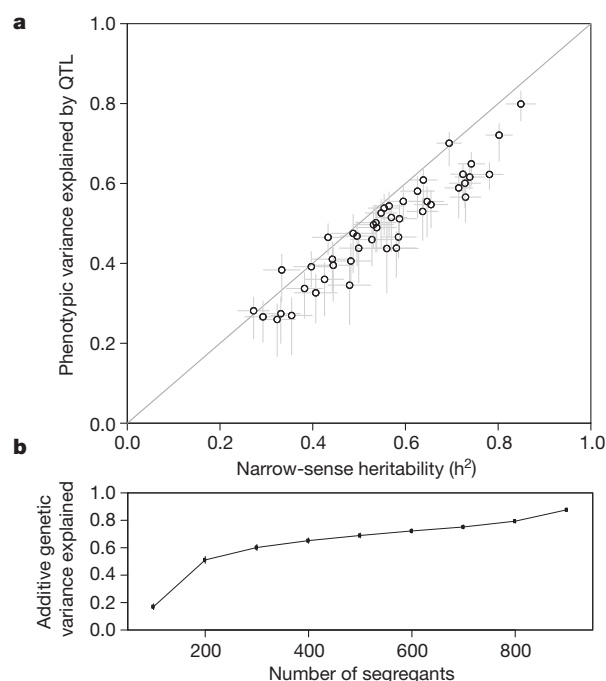
computed from dense genotype data (Methods)<sup>20</sup>. Among the 46 traits, broad-sense heritability estimates ranged from 0.40 to 0.96, with a median of 0.77. Narrow-sense heritability estimates ranged from 0.21 to 0.84, with a median of 0.52 (Fig. 1). An analysis that partitioned additive genetic variation among chromosomes produced similar results (Supplementary Table 2). We used the difference between broad-sense and narrow-sense heritability to estimate the fraction of genetic variance due to gene–gene interactions, which ranged from 0.02 to 0.54, with a median of 0.30. Thus, the genetic basis for variation in some traits is almost entirely due to additive effects, whereas for others approximately half of the heritable component is due to gene–gene interactions.

Next, we sought to map the additive heritable variation to specific quantitative trait loci (QTL). Simple linkage analysis of one marker at a time revealed multiple QTL per trait (Methods). To more accurately capture the effects of each QTL while controlling for the other QTL affecting the same trait, we used a step-wise forward-search approach to detect QTL and build a multiple-regression model (Methods). With this approach, we detected a total of 591 QTL for 46 traits at an empirical false-discovery rate (FDR) of 5% (Supplementary Table 3). We observed varying degrees of trait complexity, with a minimum of 5, a maximum of 29 and a median of 12 QTL per trait. These numbers of QTL are comparable to those previously seen for a smaller set of traits by X-QTL<sup>18</sup>. Consistent with theoretical predictions<sup>21</sup> and previous observations, we detected many more QTL of small effect than of large effect (Supplementary Fig. 3). Some traits showed a distribution of QTL effect sizes roughly consistent with Orr's evolutionary model<sup>21</sup>, whereas others showed one or more larger-than-expected QTL (Supplementary Fig. 4).

Having identified QTL, we next measured the fraction of additive heritability explained by our model of detected QTL for each trait. To obtain unbiased estimates, we performed tenfold cross-validation by detecting QTL in a subset of the segregant panel and estimating the effects in the rest of the panel. Across the traits, the detected loci explained between 72% and 100% of the narrow-sense heritability, with a median of 88% (Fig. 2a). Thus, high statistical power provided by the large segregant panel allowed us to detect QTL that jointly explain most of the additive heritability for the traits studied here. By analysing subsets of the data, we showed that 'missing' narrow-sense heritability can be explained by insufficient sample sizes (Fig. 2b). For instance, we detected 16 significant QTL, which jointly explain 78% of narrow-sense heritability for growth in E6 berbamine



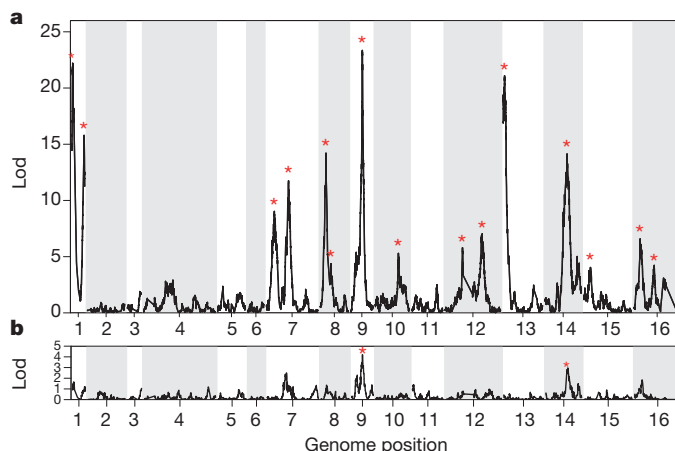
**Figure 1 | Heritability for 46 yeast traits.** The narrow-sense heritability ( $h^2$ ) for each trait is plotted against the broad-sense heritability ( $H^2$ ). Error bars show  $\pm$  s.e. in heritability estimates. The diagonal line represents  $h^2 = H^2$  and is shown as a visual guide.



**Figure 2 | Most additive heritability is explained by detected QTL.** **a**, The total variance explained by detected QTL for each trait is plotted against the narrow-sense heritability ( $h^2$ ). Error bars show  $\pm$  s.e. The diagonal line represents (variance explained by detected QTL) =  $h^2$  and is shown as a visual guide. **b**, The average fraction of additive genetic variance explained is plotted against number of segregants used for QTL detection. Error bars show  $\pm$  s.e.

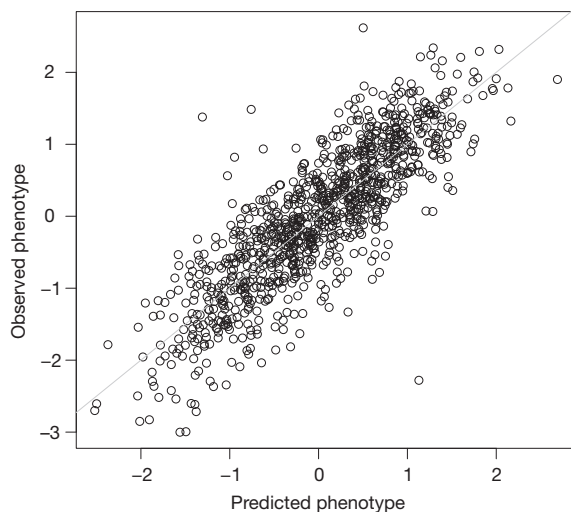
in a panel of 1,000 segregants (Fig. 3a), but only 2 of these, explaining 21% of narrow-sense heritability, also reached statistical significance in a smaller panel of 100 segregants (Fig. 3b). For traits with mostly additive genetics, the high fraction of variance explained by the detected QTL allowed us to accurately predict individual trait values from QTL genotypes (Fig. 4).

Differences between the estimates of broad-sense and narrow-sense heritability for many traits imply the presence of genetic interactions. We next sought to identify specific two-locus interactions. For each trait, we first performed an exhaustive two-dimensional scan for pairwise interactions. At a lod ( $\log_{10}$  of the odds ratio for linkage) score



**Figure 3 | QTL detection for a complex trait.** Lod score is plotted against the genetic map. Red asterisks indicate statistically significant QTL. **a**, Lod score plot with 1,005 segregants for growth in E6 berbamine. **b**, Lod score plot with 100 segregants for growth in E6 berbamine. The 15 significant QTL in **a** explain 78% of the narrow-sense heritability, compared with 21% for the 2 significant QTL in **b**. Alternating shaded bands denote chromosome boundaries.



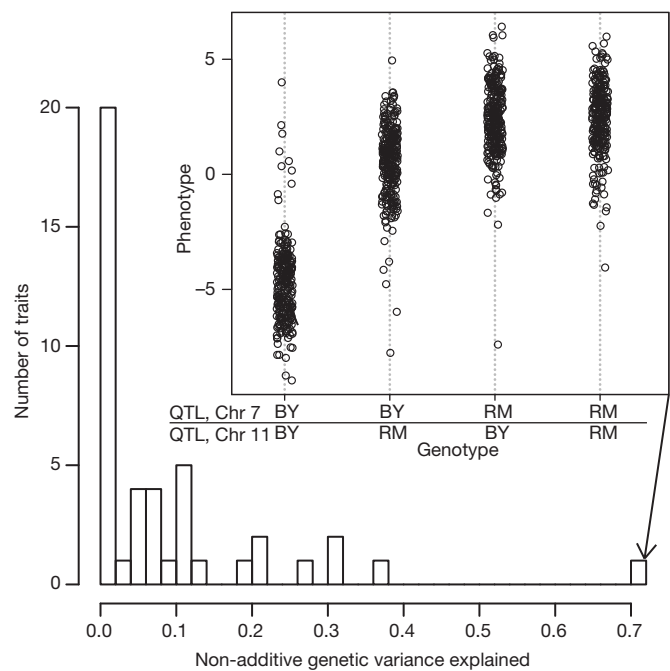


**Figure 4 | Prediction of segregant trait values from QTL phenotypes.** The observed phenotypic values for growth in lithium chloride are plotted against the predicted phenotypic values based on a cross-validated additive model of 22 QTL. The additive QTL model explains 88% of the narrow-sense heritability. The diagonal line represents (observed phenotype) = (predicted phenotype) and is shown as a visual guide.

of 6.2, corresponding to an empirical FDR of 10%, we detected significant QTL–QTL interactions for 17 of the 46 traits, with a total of 23 interacting locus pairs. A two-dimensional scan has low statistical power owing to the large search space. Power can be increased, at the cost of missing interactions between loci with no main effects, by testing only for interactions between each locus with significant additive effects and the rest of the genome<sup>22</sup>. Using this approach, we detected interactions for 24 of the 46 traits, with a total of 78 QTL–QTL interactions at an FDR of 10%. We observed a minimum of 1 and a maximum of 16 pairwise interactions per trait. These 78 pairs included 20 of the 23 locus pairs detected in the exhaustive two-dimensional scan, suggesting that two-locus interactions in which neither locus has a detectable main effect are uncommon. For 47 of the 78 pairs, both loci were detected as significant in the single-locus search for additive effects. In the remaining 31 cases, the additive effect of the second locus was too small to reach genome-wide significance, although it was nominally significant in 10 of these cases. These observations are broadly consistent with our previous work on genetic interactions that affect gene-expression traits<sup>22,23</sup>.

For most of the traits with a sizeable difference between broad-sense and narrow-sense heritability, pairwise interactions were either not detected or explained little of the difference (Fig. 5). The detected interaction effects were typically small (a median of 1.1% of genetic variance per interaction or a median of 3% of genetic variance per trait). Only in a few cases did detected genetic interactions explain a substantial fraction of the difference between broad-sense and narrow-sense heritability. Most notably, in the case of growth on maltose, one strong interaction explained 14% of the genetic variance and 71% of the difference between broad-sense and narrow-sense heritability (Fig. 5, inset).

We have used a large panel of segregants from a cross between two yeast strains to investigate the genetic architecture of 46 quantitative traits. We measured both the total and the additive contributions of genetic factors to trait variation, and showed that these often differ. The observed differences between total and additive heritability estimates suggest that the contribution of genetic interactions to broad-sense heritability ranges from near zero to 54%. However, with a few exceptions, the specific combinations of loci that account for these interactions remain elusive. There are several possible explanations for this result. First, the statistical power to detect interactions is lower than the power to detect main effects. Second, individual interaction



**Figure 5 | Non-additive genetic variance explained by QTL–QTL interactions.** A histogram of the fraction of non-additive genetic variance explained by detected QTL–QTL interactions per trait is plotted. The histogram is restricted to traits for which at least 10% of the total genetic variation is non-additive. Inset, phenotypes for growth in maltose are shown, grouped by two-locus genotypes at the two interacting QTL on chromosomes 7 and 11. This QTL–QTL interaction explained 71% of the difference between broad-sense and narrow-sense heritability.

effects are expected to be smaller than additive effects, and hence their detection requires even larger sample sizes<sup>13</sup>. Finally, higher-order interactions among more than two loci could also contribute<sup>24</sup>. Our estimates of the contribution of interactions in a cross may overestimate their contribution to trait heritability in a population, because a higher proportion of variance is expected to be additive as allele frequencies depart from one-half<sup>25</sup>.

The large size of the panel allowed us to detect specific loci that jointly account for almost all of the additive (narrow-sense) heritability of each trait (72–100%). Human traits examined by GWAS vary in their genetic complexity<sup>1</sup>, ranging from macular degeneration, for which 5 variants in 3 genes explain roughly half of the genetic risk<sup>26</sup>, to height, for which 180 loci explain about 13% of heritability, implying the existence of a much larger number of undetected loci<sup>27</sup>. Compared to our results in a yeast cross, GWAS typically detect a larger number of loci explaining a smaller proportion of trait heritability. One obvious difference is that the number of variants segregating in a cross between two strains is smaller than the number of common variants segregating in a population sample. The difference is roughly a factor of three for a neutral allele frequency spectrum, and potentially much larger if functional variants are deleterious and hence shifted towards lower frequency (Methods). The human genome also offers a larger target size, perhaps by a factor of five, for variants affecting a trait (Methods). These very rough estimates suggest that we might expect at least 15 times more loci to be found by GWAS than the median of 12 loci per trait we observe in the yeast cross. Because of the resolution of linkage analysis in our cross, some QTL may contain multiple linked variants, further increasing the true number of loci. Several additional factors could lead to a larger missing heritability in humans: the fraction of heritability due to genetic interactions could be higher<sup>13</sup>, rare variants may account for a disproportionately large contribution of heritable variation<sup>28–30</sup>, and some human traits might be inherently more complex than yeast traits in that they integrate over physiological

processes involving a larger number of underlying gene pathways. Within-locus dominance effects represent an additional source of genetic complexity in diploid organisms.

Our results are consistent with the suggestions that missing additive (narrow-sense) heritability arises primarily from many loci with small but not infinitesimal effects. These loci can be discovered in studies with sufficiently large sample sizes, although the optimal study designs will depend on the population frequency spectra of the causative alleles. Because all alleles are fixed at a frequency of one-half in a cross, we cannot yet delineate the contributions of common and rare variants to inherited variation, but we plan to do so in future studies.

## METHODS SUMMARY

**Construction, genotyping and phenotyping of segregant panel.** We crossed two prototrophic strains, a *MATa* BY parent (a laboratory strain derived from a cross of BY4716 and BY4700) and a *MAT $\alpha$*  RM parent (a vineyard strain derived from RM11-1a). We sporulated the diploid and retained 1,056 four-spore tetrads that showed 2:2 segregation of mating type and drug-resistance markers. One spore from each tetrad was genotyped using a modified version of the Nextera protocol and Illumina short-read sequencing. Segregants were phenotyped for end-point growth on agar plates. Custom image processing software was used to quantify colony size.

**Statistical analysis.** We estimated broad-sense heritability using replicated trait measurements for each strain and a random effects analysis of variance. We estimated narrow-sense heritability using a linear mixed model that compares phenotypic similarity among individuals with their realized genetic relatedness measured from genotype data. Standard methods were used for linkage analysis. Empirical FDRs were calculated using a permutation approach.

**Full Methods** and any associated references are available in the online version of the paper.

Received 27 June; accepted 14 December 2012.

Published online 3 February 2013.

- Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
- Hill, W. G. Understanding and using quantitative genetic variation. *Phil. Trans. R. Soc. Lond. B* **365**, 73–85 (2010).
- Mackay, T. F. C., Stone, E. A. & Ayroles, J. F. The genetics of quantitative traits: challenges and prospects. *Nature Rev. Genet.* **10**, 565–577 (2009).
- Buckler, E. S. *et al.* The genetic architecture of maize flowering time. *Science* **325**, 714–718 (2009).
- Atwell, S. *et al.* Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627–631 (2010).
- Mackay, T. F. *et al.* The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**, 173–178 (2012).
- Aylor, D. L. *et al.* Genetic analysis of complex traits in the emerging Collaborative Cross. *Genome Res.* **21**, 1213–1222 (2011).
- Rockman, M. V. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66**, 1–17 (2012).
- Goldstein, D. B. Common genetic variation and human traits. *N. Engl. J. Med.* **360**, 1696–1698 (2009).
- Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genet.* **42**, 565–569 (2010).
- Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
- Pritchard, J. K. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* **69**, 124–137 (2001).
- Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl Acad. Sci. USA* **109**, 1193–1198 (2012).
- Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* edn 4 (Longman, 1996).
- Eichler, E. E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Rev. Genet.* **11**, 446–450 (2010).
- Ehrenreich, I. M., Gerke, J. P. & Kruglyak, L. Genetic dissection of complex traits in yeast: insights from studies of gene expression and other phenotypes in the BY  $\times$  RM cross. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 145–153 (2009).
- Brem, R. B. & Kruglyak, L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl Acad. Sci. USA* **102**, 1572–1577 (2005).
- Ehrenreich, I. M. *et al.* Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464**, 1039–1042 (2010).
- Ruderfer, D. M., Pratt, S. C., Seidel, H. S. & Kruglyak, L. Population genomic analysis of outcrossing and recombination in yeast. *Nature Genet.* **38**, 1077–1081 (2006).
- Visscher, P. M. *et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* **2**, e41 (2006).
- Orr, H. A. Adaptation and the cost of complexity. *Evolution* **54**, 13–20 (2000).
- Storey, J. D., Akey, J. M. & Kruglyak, L. Multiple locus linkage analysis of genomewide expression in yeast. *PLoS Biol.* **3**, e267 (2005).
- Brem, R. B., Storey, J. D., Whittle, J. & Kruglyak, L. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* **436**, 701–703 (2005).
- Dowell, R. D. *et al.* Genotype to phenotype: a complex problem. *Science* **328**, 469 (2010).
- Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4**, e1000008 (2008).
- Maller, J. *et al.* Common variation in three genes, including a noncoding variant in *CFH*, strongly influences risk of age-related macular degeneration. *Nature Genet.* **38**, 1055–1059 (2006).
- Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
- Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
- Keinan, A. & Clark, A. G. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* **336**, 740–743 (2012).
- Nelson, M. R. *et al.* An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337**, 100–104 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank D. Botstein, M. McClean, E. Andersen, F. Albert, S. Treusch, R. Ghosh and X. Wang for comments on the manuscript, Y. Jia and S. Schrader for technical assistance and E. Lander for discussions. This work was supported by National Institutes of Health (NIH) grants R37 MH59520 and R01 GM102308, a James S. McDonnell Centennial Fellowship, and the Howard Hughes Medical Institute (L.K.), a National Science Foundation (NSF) fellowship (J.S.B.), NIH postdoctoral fellowship F32 HG51762 (I.M.E.) and NIH grant P50 GM071508 to the Center for Quantitative Biology at the Lewis-Sigler Institute of Princeton University.

**Author Contributions** Experiments were designed by J.S.B., I.M.E. and L.K. Experiments were performed by J.S.B., I.M.E., W.T.L. and T.-L.V.L. Analyses were conducted by J.S.B. The manuscript was written by J.S.B. and L.K. and incorporates comments by all other authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.K. ([leonid@genomics.princeton.edu](mailto:leonid@genomics.princeton.edu)).

## METHODS

**Construction of segregant panel.** We crossed a prototrophic BY parent (a laboratory strain derived from a cross of BY4716 and BY4700) that is *MATa* with a prototrophic RM parent (a vineyard strain derived from RM11-1a) that is *MATα hoA::hphMX4 flo8Δ::natMX4 AMN1-BY*. Diploid zygotes were recovered and sporulated for 5–7 days in SPO++ sporulation medium (<http://dunham.gs.washington.edu/sporulationdissection.htm>) in a roller drum at room temperature (~21–22 °C). Tetrads were dissected<sup>31</sup> using the MSM 400 dissection microscope (Singer Instrument Company). Colonies from four-spore tetrads were inoculated into 150 µl of yeast nitrogen base (YNB) + 2% glucose in 96-well plates (Corning), grown for 48 h at 30 °C without shaking, and stored as frozen stocks in 20% glycerol. 1,056 four-spore tetrads that showed 2:2 segregation of hygromycin resistance, G418 resistance and mating type were retained. A Biomek FX (Beckman Coulter) was used to select one segregant from each four-spore tetrad for downstream analysis.

**Power calculations.** We calculated statistical power ( $1 - \beta$ ) for sample sizes of 100 and 1,000 segregants in R using the 'power.t.test' function<sup>32</sup>. Power was calculated over a range of effect sizes, where effect size was calculated as the per cent genetic variance explained by a single QTL. To correct for multiple testing over thousands of markers across the genome, a genome-wide significance threshold ( $\alpha$ ) of  $P < 2 \times 10^{-4}$  was used.

**DNA preparation and sequencing library construction.** Segregants were inoculated into 1-ml deep-well 96-well plates (Thermo Scientific) in 800 µl of yeast peptone dextrose (YPD) and grown 2 days at 30 °C without shaking. Plates were sealed with Breathe-Easy gas-permeable membranes (Sigma-Aldrich). DNA was extracted using 96-well DNeasy Blood & Tissue kits (Qiagen). DNA concentrations were determined using the Quant-iT dsDNA High-Sensitivity DNA quantification kit (Invitrogen) and the Bio-Tek Synergy 2 plate reader. DNA was diluted to 1.66 ng per microlitre. Per sample, 15 µl of 1.66 ng per µl DNA was added to 4 µl of 5× Nextera HMW buffer, 0.95 µl of water and 0.05 µl of Nextera Enzyme Mix (Epicentre). The transposition reaction was performed for 5 min at 55 °C. One-hundred microlitres of water was added to each sample, and the samples were purified using the MinElute kit (Qiagen). Fifteen microlitres of the purified fragmented DNA was PCR-amplified and barcoded with custom 5-base pair (bp) sequences using Ex Taq polymerase (Takara) and between 15 and 30 PCR cycles. Five microlitres of each PCR-amplified sample was combined into one 96-plex library. The combined library was loaded on a 2% agarose gel, and the 350–650-bp region was excised and gel-extracted using QIAquick Gel Extraction Kit (Qiagen). Final libraries were diluted to 3.3 ng per microlitre and sequenced using the single-end module on a HiSeq 2000 (Illumina) with 100-bp reads.

**Determining segregant genotypes.** Custom R and python code was used to demultiplex the sequencing data and trim ends. Sequencing reads were assigned to segregants based on the 5-bp barcode at the beginning of each read. The internal 19-bp transposon sequence and 10 bp on the right end of each read were removed. Reads were aligned to the S288C reference genome using the Burrows–Wheeler Aligner<sup>33</sup> with the '-q 30' parameter. SAMtools<sup>34</sup> was run with the 'view' command and '-bHsq 1' parameters to retain uniquely mapping reads. Sequence variants were identified using SAMtools<sup>34</sup> with the 'mpileup' command and parameters '-d 10000 -D -u'. 42,689 high-confidence sequence variants between BY and RM were determined from sequencing the parental strains at greater than 50-fold coverage. Variants in the segregants were restricted to these 42,689 expected sites using 'bcftools view'<sup>34</sup> with the parameters '-N -c -g -v -P flat'. Genotype likelihoods for the BY and RM alleles for each genotypic variant were extracted from the VCF file using custom R code.

For each segregant and chromosome, a hidden Markov model was used to calculate the posterior likelihood that the read data was coming from the BY allele, the RM allele, a BY gene-conversion event or an RM gene-conversion event. Genotypes were called the BY variant if the  $\log_{10}$  ratio of the BY posterior likelihood and the RM posterior likelihood was greater than 2, the RM variant if this ratio was less than -2, and missing data if between -2 and 2. In total, 1,008 out of 1,056 segregants had between 25 and 120 recombination breakpoints and at least 35,000 markers with genotype calls, and these were retained for downstream analysis. Genotypic markers were excluded if their allele frequency was greater than 56% or less than 45%, or if they were not called in at least 99% of the segregants. This resulted in a final set of 30,594 genotypic markers. Markers with missing data were imputed using the Viterbi algorithm as implemented in the R/qtl package<sup>35</sup>. Adjacent markers with the same genotypes in all segregants were collapsed to one unique marker, resulting in a final set of 11,623 unique genotypic markers.

**Phenotyping by end-point growth on agar plates.** Individual segregants were inoculated in at least two different plate configurations into 384-well plates (Thermo Scientific; 264574) with 50 ml of YPD and grown for 36 to 48 h in a 30 °C incubator without shaking. Replicates here consisted of strains that were

grown in independent 384-well plates and in different plate configurations. Each segregant was represented once in each plate configuration. Each target agar plate was made with 50 ml of media (YPD or YNB) and with drug or condition of choice (Supplementary Table 1). The Singer Rotor HDA pinning robot was used to pin the segregants to the agar plates. Before pinning, each 384-well plate was mixed for 1 min at 2,000 r.p.m. using a MixMate (Eppendorf). Segregants were pinned to the agar plates with 100% pin pressure and 384 long pins (Singer Instruments; RP-MP-3L). After pinning, the plates were incubated at 30 °C, or the specified condition temperature (Supplementary Table 1), for approximately 48 h. Plates were scanned face-up and without lids on an Epson 700 transparency scanner with 400 DPI resolution and a greyscale bit-depth of 8. Images were saved as TIFFs or 99% quality JPEGs. The pixel coordinates of the centres of the four corner colonies for each plate were manually identified using ImageJ<sup>36</sup>.

Custom R code was written to determine the size of each colony. Expected colony positions were calculated using manually identified coordinates of the corner colonies for each plate. Images were segmented using k-means clustering on the distribution of pixel intensities across a plate. The Voronoi region of each 9-pixel diameter circular seed, corresponding to the expected location of a colony, was used together with the segmented image to match colonies to their expected positions. This was implemented using functions in the EBIImage<sup>37</sup> R package. The

radius of each colony was calculated as  $\sqrt{\frac{\text{pixelcount}}{\pi}}$ . Colonies with more than 15 pixels touching the edge of the image and colonies larger than 3,500 pixels were removed and treated as missing data in downstream analysis. Irregular colonies, representing image processing or pinning artefacts, were removed if  $\frac{\text{perimeter}}{2\sqrt{\text{pixelcount}}} > 1$  and radius  $> 20$ . These irregular colonies were also treated as missing data for downstream analysis. Colonies on each edge of a plate were tested for difference with all other colonies on the plate using a Wilcoxon rank-sum test. All colonies on an edge with  $P < 0.05$  were treated as missing data and excluded from downstream analysis. Images were further inspected manually, and colonies subject to pinning or image-processing artefacts were removed. To normalize for occasional subtle within-plate spatial growth artefacts, a robust locally weighted regression was fit to the radius measurements using functions in the locfit R package<sup>38</sup>. The residuals were used for downstream analysis. End-point growth measurements were normalized for growth on control media by fitting a regression for effect of growth on control media and using the residuals for downstream analysis. One-hundred-and-forty conditions were assayed. If multiple doses of a compound were tested, the dose with the highest heritability and phenotype data for at least 600 segregants was retained. Traits with a narrow-sense heritability of less than 25% were excluded from downstream analysis.

**Calculating heritability.** Broad-sense heritability was calculated using replicated segregant data and a random effects analysis of variance. This was implemented using the 'lmer' function in the lme4 R package<sup>39</sup>. The variance components  $\sigma_G^2$ , the genetic variance due to effect of segregant; and  $\sigma_E^2$ , the error variance, were calculated, and broad sense-heritability was estimated as  $\frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2}$ . Standard errors were calculated by delete-one jackknife.

Narrow-sense heritability was calculated for each trait using a linear mixed model<sup>40</sup>. This can be written as  $y = \beta 1_n + Zu + e$ . Here  $y$  is a vector of phenotype values for  $n$  segregants. For the comparison of broad-sense and narrow-sense heritabilities,  $y$  consists of one randomly chosen measurement of the replicate phenotype measurements for each segregant. This puts the comparison of broad-sense and narrow-sense heritability on the same scale. For the comparison of variance explained by each QTL model and narrow-sense heritability,  $y$  consists of the average of replicate phenotype measurements for each segregant.  $\beta$  is the overall mean, and  $1_n$  is a vector of  $n$  ones.  $Z$  is an  $n \times n$  identity matrix,  $u$  is a vector of random effects (BLUPs or breeding values for each segregant), and  $e$  is a vector of residuals. The variance structure of the phenotypes is written as  $V = A\sigma_A^2 + I\sigma_{EV}^2$ , where  $A$  is the relatedness matrix between all pairs of segregants, estimated from our genotype data as the proportion of markers shared identical-by descent (IBD) between all pair of segregants,  $I$  is an  $n \times n$  identity matrix,  $\sigma_A^2$  is the polygenic additive genetic variance explained by the single nucleotide polymorphisms and  $\sigma_{EV}^2$  is the error variance. Variance components were estimated using the rrBLUP R package<sup>41</sup>, and narrow-sense heritability was estimated as  $\frac{\sigma_A^2}{\sigma_A^2 + \sigma_{EV}^2}$ . Standard errors were calculated by delete-one jackknife. Estimates of narrow-sense heritability using the average of segregant replicates or one measurement for each segregant were similar.

Although the average genetic relatedness among the segregants is 0.5, it varies owing to random Mendelian segregation, and the actual relatedness for any pair of segregants can be calculated from high-density genotype data as the proportion of single nucleotide polymorphism alleles shared by these segregants. In our segregant panel, the standard deviation of relatedness was 0.048. For the partitioning of



additive variance by chromosome, the relatedness matrix  $A$  between all pairs of segregants was estimated as the proportion of markers on a chromosome-shared IBD between all pair of segregants<sup>42</sup>.

**Mapping additive QTL.** Forty-six traits were chosen for QTL mapping on the basis of the criteria described above. Each trait was scaled to have mean 0 and variance 1. We tested for linkage by calculating lod scores for each genotypic marker and each trait as  $-n(\ln(1-r^2)/2\ln(10))$ , where  $r$  is the Pearson correlation coefficient between the segregant genotypes at the marker and segregant trait values<sup>43</sup>.

To estimate significance empirically, assignment of phenotype to each segregant was randomly permuted 1,000 times while maintaining the correlation structure among phenotypes. The maximum lod score for each chromosome and trait was retained<sup>44</sup>. The FDR was calculated as the ratio of expected peaks to observed peaks across different lod thresholds. Genetic markers corresponding to QTL peaks that were significant at an FDR of 5% were added to a linear model for each trait. Trait-specific linear models that included the significant QTL genotypes as additive covariates were computed, and phenotypic residuals were estimated. Phenotypic residuals for each trait were then used for another round of QTL detection<sup>45</sup>. This process of peak detection, calculation of empirical significance thresholds and expansion of the linear model for each trait to include significant QTLs detected at each step was repeated four times. The lod thresholds corresponding to a 5% FDR at each step were 2.68, 2.92, 3.72 and 4.9.

**Calculating effect size and variance explained by additive QTL.** For each trait, a multiple regression linear model was computed with trait-specific QTL genotypes as independent variables. Phenotypes for each trait were scaled to have mean 0 and variance 1. The multiple regression coefficients are the standardized differences in allelic means for each QTL while controlling for the effects of other segregating QTL. The square of the multiple regression coefficient is the fraction of phenotypic variance explained by a QTL. The fitted truncated exponential distribution in Supplementary Fig. 3 is parameterized as  $\frac{bnw}{e^{-lb} + e^{-rb}} e^{-bx}$ , where  $x$  is the absolute value of the multiple regression coefficient pooled across traits and QTL,  $n$  is the number of bins,  $w$  is the bin size,  $l$  and  $r$  are left and right truncation points, respectively and  $b$  is estimated using maximum likelihood. Here  $l = 0.12$  and  $r = 0.35$  to correspond to the magnitude of effect sizes where power is nearly 100% and to exclude large-effect QTL.

The total phenotypic variance explained by the multiple regression QTL model is the  $r^2$  from the model, which was calculated using the 'fitqtl' function in R/qtl. Unbiased estimates of the total phenotypic variance explained by the multiple QTL model were calculated by standard tenfold cross validation. The segregants were randomly split into ten equal-sized groups, nine groups were combined for QTL detection using the algorithm described above, and the remaining group was used to estimate QTL effect sizes. This process was repeated for each of the ten groups, and the average of the ten estimates was calculated. A similar method was used for phenotypic prediction. The segregants were randomly split into ten equal-sized groups, but here nine groups were combined for QTL detection and for estimation of QTL effect sizes. Phenotypes were then predicted for the remaining group (validation set) using only the genotypes from the validation set and the QTL model constructed from the nine other groups (training set). This was repeated such that each segregant was in the validation set one time and the training set nine times.

**Detecting QTL-QTL interactions.** For computational efficiency, the marker set was reduced to 4,420 by picking one marker closest to each centimorgan position on the genetic map. For the full-genome scan for interacting QTL, a lod score corresponding to the likelihood ratio of a model that includes an interaction term,  $y = ax + bz + cxz + d$ , to a model that does not,  $y = ax + bz + d$ , was computed for each trait and every marker pair. Here,  $y$  is the residuals vector for each trait after fitting the additive QTL model,  $x$  is the genotype vector at one position in the genome,  $z$  is the genotype vector at another position in the genome at least 25 centimorgans away from  $x$ , and  $a$ ,  $b$ ,  $c$  and  $d$  are estimated parameters specific to each trait and marker pair<sup>22,23</sup>. FDR at different lod thresholds was calculated by dividing the average number of peaks obtained in 1,000 permutations of the data that scramble the segregant identities by the number of peaks observed in the real data.

To increase statistical power, we tested for interactions between each locus with significant additive effects and the rest of the genome. A lod score for interaction was computed in exactly the same manner as for the full two-dimensional scan described above, except that  $x$  was constrained to genotypes corresponding to trait-specific significant additive QTL. FDR was calculated as above. Non-additive genetic variance explained was calculated as the amount of variance explained by including QTL-QTL interactions in the model divided by the difference between broad-sense and narrow-sense heritability.

**Conversion factors for yeast versus human expected numbers of QTL.** The number of variants segregating in a cross between two strains is smaller than the number of (common) variants segregating in a population sample. If the causal variants follow a neutral allele frequency spectrum, then under the standard neutral model of population genetics, the relationship between the number of variants segregating in a cross of two haploid strains ( $S_2$ ) and the number of common variants segregating in a population ( $S_f$ ) with minor allele frequency  $\geq f$  is  $S_f = S_2 \ln(\frac{1-f}{f})$  (ref. 46). By setting  $f$  to 0.05, we obtain a ratio  $S_f/S_2$  of approximately three (2.94). The ratio is 2.2 for  $f = 0.1$  and 4.6 for  $f = 0.01$ . If the functional variants are at least weakly deleterious, and hence skewed towards lower frequency, the conversion factors will be larger.

Several approaches can be used to obtain a rough idea of relative target size for mutation. The yeast genome is 12 megabases, whereas the coding regions of the human genome alone are  $\sim 30$  megabases, and at least the same amount of non-coding sequence is expected to be functional based on selective constraint<sup>47</sup>. If we assume that the number of loci involved in a trait is proportional to the number of bases available for mutations with functional consequences, human traits should be more complex than yeast traits by at least a factor of 5. Similarly, yeast has  $\sim 5,700$  genes compared to  $\sim 20,000$  for humans, which gives a conversion factor of  $\sim 3.5$ .

31. Amberg, D. C., Burke, D. & Strathern, J. N. *Methods in Yeast Genetics: a Cold Spring Harbor Laboratory Course Manual* (Cold Spring Harbor Laboratory Press, 2005).
32. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2012).
33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
34. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
35. Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
36. Abramoff, M. D., Magalhaes, P. J. & Ram, S. J. Image Processing with ImageJ. *Biophotonics International* **11**, 36–42 (2004).
37. Pau, G., Fuchs, F., Sklyar, O., Boutros, M. & Huber, W. EImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics* **26**, 979–981 (2010).
38. Loader, C. *locfit: Local Regression, Likelihood and Density Estimation* <http://CRAN.R-project.org/package=locfit> (2012).
39. Bates, D., Maechler, M. & Bolker, B. *lme4: Linear Mixed-Effects Models Using Eigen and S4* <http://CRAN.R-project.org/package=lme4> (2011).
40. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).
41. Endelman, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* **4**, 250–255 (2011).
42. Visscher, P. M. Variation of estimates of SNP and haplotype diversity and linkage disequilibrium in samples from the same population due to experimental and evolutionary sample size. *Ann. Hum. Genet.* **71**, 119–126 (2007).
43. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* edn 1 (Sinauer Associates, 1998).
44. Chen, L. & Storey, J. D. Relaxed significance criteria for linkage analysis. *Genetics* **173**, 2371–2381 (2006).
45. Doerge, R. W. & Churchill, G. A. Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**, 285–294 (1996).
46. Kruglyak, L. & Nickerson, D. A. Variation is the spice of life. *Nature Genet.* **27**, 234–236 (2001).
47. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).

# Concurrent activation of striatal direct and indirect pathways during action initiation

Guohong Cui<sup>1\*</sup>, Sang Beom Jun<sup>2\*</sup>, Xin Jin<sup>1,3</sup>, Michael D. Pham<sup>1</sup>, Steven S. Vogel<sup>4</sup>, David M. Lovinger<sup>1,5</sup> & Rui M. Costa<sup>1,6</sup>

The basal ganglia are subcortical nuclei that control voluntary actions, and they are affected by a number of debilitating neurological disorders<sup>1–4</sup>. The prevailing model of basal ganglia function proposes that two orthogonal projection circuits originating from distinct populations of spiny projection neurons (SPNs) in the striatum<sup>5,6</sup>—the so-called direct and indirect pathways—have opposing effects on movement: activity of direct-pathway SPNs is thought to facilitate movement, whereas activity of indirect-pathway SPNs is presumed to inhibit movement<sup>1,2</sup>. This model has been difficult to test owing to the lack of methods to selectively measure the activity of direct- and indirect-pathway SPNs in freely moving animals. Here we develop a novel *in vivo* method to specifically measure direct- and indirect-pathway SPN activity, using Cre-dependent viral expression of the genetically encoded calcium indicator (GECI) GCaMP3 in the dorsal striatum of D1-Cre (direct-pathway-specific<sup>6,7</sup>) and A2A-Cre (indirect-pathway-specific<sup>8,9</sup>) mice<sup>10</sup>. Using fibre optics and time-correlated single-photon counting (TCSPC) in mice performing an operant task, we observed transient increases in neural activity in both direct- and indirect-pathway SPNs when animals initiated actions, but not when they were inactive. Concurrent activation of SPNs from both pathways in one hemisphere preceded the initiation of contraversive movements and predicted the occurrence of specific movements within 500 ms. These observations challenge the classical view of basal ganglia function and may have implications for understanding the origin of motor symptoms in basal ganglia disorders.

Excitatory inputs to striatum arise from cortical and thalamic structures, while striatal inhibitory output diverges, with some neurons projecting directly to basal ganglia output nuclei (striatonigral SPNs) and others to intermediate nuclei (striatopallidal SPNs), giving rise to the direct and the indirect pathways that propagate throughout the basal ganglia<sup>6,7</sup>. Despite advances in our understanding of basal ganglia circuitry, the functional relationship between these two pathways and the generation of actions is still under debate. Although the predominant model proposes an ‘opposing’ scheme primarily based on the polarity of neural transmission in these two pathways and their final convergence onto basal ganglia output nuclei<sup>2</sup>, more recent models propose coordinated activation of both pathways during action selection. Some propose, for example, that activation of the direct pathway could facilitate output of the desired motor programs but that activation of the indirect pathway would inhibit competing motor programs<sup>3,4,11</sup>. Others suggest that coordinated activity of direct and indirect pathway is critical for the appropriate timing and synchrony of basal ganglia circuits during movement<sup>12,13</sup>.

Disambiguating between these models requires monitoring and comparing the activity of direct- and indirect-pathway SPNs in behaving animals<sup>4,11</sup>. However, electrophysiological recordings have not been useful for this purpose because there are no obvious electrophysiological

parameters to distinguish between direct- and indirect-pathway SPN firing, and antidromic stimulation can be difficult<sup>14</sup> and unreliable considering the crosstalk between axon collaterals of the two pathways<sup>14,15</sup>. Optogenetic techniques have permitted the identification of specific cell types *in vivo* through optical stimulation of recorded cells<sup>16</sup> or optical monitoring of the activity of cells of a specific cell type<sup>17</sup>. This latter approach has permitted the measurement of neural-activity-dependent fluorescence changes from specific types of neurons in behaving animals<sup>17–19</sup>. However, current *in vivo* imaging methods are not ideal for exploring the activity of subcortical structures in freely moving animals owing to limitations in the penetration depth (<1 mm) and the general requirement for head restraint during measurement<sup>20</sup>.

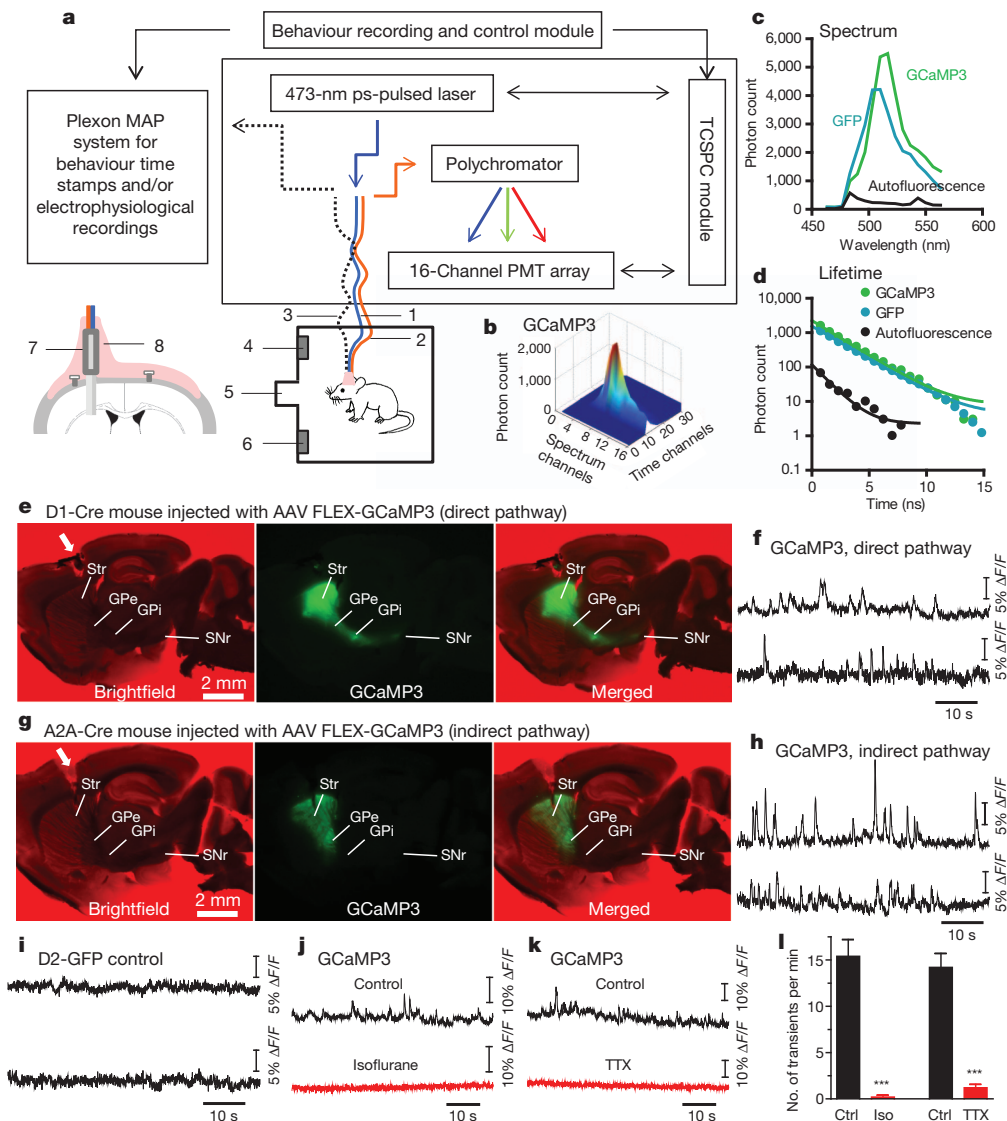
To circumvent these issues and achieve optical recording of neural activity from striatum in freely moving animals, we developed an *in vivo* photometry method using time-correlated single-photon counting (TCSPC)-based fibre optics. The TCSPC system includes a 473-nm picosecond-pulsed laser for excitation, a polychromator to disperse the fluorescent emission, a 16-channel photomultiplier (PMT) array for photon detection and a TCSPC module for single-photon counting (Fig. 1a). A fibre-optic probe consisting of a single-mode fibre (used to deliver laser pulses) and a multi-mode fibre (to collect emitted photons) is implanted into the desired structure in mouse brain; in our case the dorsal striatum. Photons collected by the TCSPC system can be plotted as time-resolved spectra (Fig. 1b). These plots are used to determine the emission spectrum, the fluorescence lifetime of each spectral component, and the integrated fluorescence intensity (Fig. 1c, d). Because the emission spectrum and lifetime of a fluorescent indicator are often different from background autofluorescence, these characteristics can be used to discriminate between these signals. As GCaMP3 is an intensity-based GECI whose lifetime does not change with Ca<sup>2+</sup> binding (Supplementary Fig. 1), we used changes in GCaMP3 fluorescence intensity (‘fluorescent transients’) as a readout of neural activity throughout this study, and lifetime and emission spectra as controls for non-specific fluorescence.

The single-mode and multi-mode fibres were joined at the free ends to form a parallel hybrid fibre probe, which was inserted into the dorsal striatum (Fig. 1a and Supplementary Fig. 2a, b), and were light and flexible enough to allow animals to move freely (Supplementary Video 1). The effective detection volume of this hybrid probe is the overlap between the volume excited by the single-mode fibre and the acceptance cone of the multi-mode fibre. In fluorescent brain tissue this volume had a depth of around 500 µm (see Methods and Supplementary Fig. 2c–f).

To express GCaMP3 selectively in direct- and indirect-pathway SPNs, we injected recombinant adeno-associated virus vectors (AAV2/9) containing a double-floxed inverted open reading frame encoding GCaMP3 (FLEX-GCaMP3 (flip-excision genetic switch)) unilaterally into the left dorsal striatum of D1-Cre and A2A-Cre

<sup>1</sup>Section on In Vivo Neural Function, Laboratory for Integrative Neuroscience, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, 5625 Fishers Lane, Bethesda, Maryland 20892-9412, USA. <sup>2</sup>Department of Electronics Engineering, Ewha Womans University, Seoul 120-750, Korea. <sup>3</sup>Molecular Neurobiology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. <sup>4</sup>Section on Cellular Biophotonics, Laboratory for Molecular Physiology, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, 5625 Fishers Lane, Bethesda 20892-9412, Maryland, USA. <sup>5</sup>Section on Synaptic Pharmacology, Laboratory for Integrative Neuroscience, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, 5625 Fishers Lane, Bethesda, Maryland 20892-9412, USA. <sup>6</sup>Champalimaud Neuroscience Programme at Instituto Gulbenkian de Ciência and Champalimaud Centre for the Unknown, Lisbon 1400-038, Portugal.

\*These authors contributed equally to this work.



**Figure 1 | Optical measurement of neural-activity-dependent fluorescence changes in direct- and indirect-pathway SPNs in freely moving mice.**

**a**, Schematic of TCSPC-based fibre-optic system, showing the single-mode fibre for excitation (1), multi-mode fibre for photon collection (2), optional electrophysiological cable (3), right lever (4), food magazine (5), and left lever (6). The inset (lower left part) shows the hybrid fibre probe (7) lowered into dorsal striatum and fixed in place by dental acrylic (8). **b**, Time-resolved spectrum of GCaMP3 expressed in striatal SPNs. **c**, **d**, Fluorescence spectrum (c) and lifetime (d) of GCaMP3, GFP and brain autofluorescence measured from dorsal striatum of GCaMP3-expressing A2A-Cre, D2-GFP and wild-type mice, respectively. **e–h**, GCaMP3 fluorescence from acutely cut brain slices

mice<sup>10</sup>. GCaMP3 was therefore expressed specifically in direct-pathway SPNs (axons projecting to the two main output nuclei of the basal ganglia: the internal segment of globus pallidus (GPi) and substantia nigra pars reticulata (SNr), Fig. 1e) in D1-Cre mice<sup>21</sup> or in indirect-pathway SPNs (axons terminating in the external segment of globus pallidus (GPe) (Fig. 1g), and specific expression in indirect-pathway SPNs (Supplementary Fig. 3)) in A2A-Cre mice.

In all freely moving D1-Cre and A2A-Cre mice with a distinct GCaMP3 spectrum, we routinely observed transient increases in fluorescence intensity (Fig. 1f, h and Supplementary Fig. 5), resembling the locomotor-related transients measured previously with microscopy in motor cortex neurons<sup>18</sup>. Transients were more prominent on spectral channels that detect the peak of the GCaMP3 emission spectrum than on off-peak spectral channels in GCaMP3-expressing mice

showing GCaMP3 selectively expressed in direct- and indirect-pathway SPNs in D1-Cre (e) and A2A-Cre (g) mice, respectively (white arrowheads indicate the fibre-probe tract); and examples of striatal GCaMP3 fluorescence recorded in two freely moving D1-Cre (f) and A2A-Cre (h) mice. **i**, Examples of striatal GFP fluorescence recorded in two freely moving D2-GFP mice. **j**, **k**, Example traces showing that isoﬂurane anaesthesia (**j**) and intra-striatal TTX injection (**k**) abolished GCaMP3 transients in A2A-Cre mice. **l**, Summary of isoﬂurane (Iso) and TTX effects on GCaMP3 transient frequency. \*\*\* $P < 0.001$ , unpaired  $t$ -test,  $n = 5$  bins under each condition per animal, bin size = 1 min. Error bars indicate + s.e.m. Ctrl, control; Str, striatum.

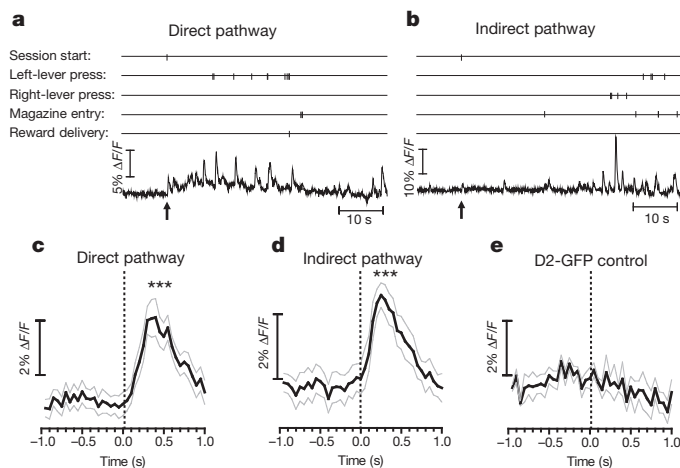
(Supplementary Fig. 6), and were not seen in wild-type mice (Supplementary Fig. 7) or mice that expressed green fluorescent protein (GFP) (Fig. 1i and Supplementary Fig. 5). Thus, transients are unlikely to reflect motion-related artefacts. GCaMP3 fluorescence transients decreased substantially after intra-striatal injection of tetrodotoxin (TTX, 200 nM, 500 nl, Fig. 1k, l and Supplementary Fig. 9), supporting the idea that they reflect neuronal activity. Consistent with this, deep isoﬂurane anaesthesia, which temporarily but markedly diminishes activity in the striatum (Supplementary Fig. 8), completely and reversibly blocked GCaMP3 transients (Fig. 1j, l) and caused a drop in baseline fluorescence (Supplementary Fig. 7). These effects were occluded by previous TTX administration, indicating that isoﬂurane-induced loss of transients reflected decreased neural activity (Supplementary Fig. 9).



To determine the relationship between the GCaMP3 transients and SPN firing patterns, we carried out  $\text{Ca}^{2+}$  imaging and cell-attached recordings in acutely cut striatal slices. We evoked action potentials by current injection through the patch pipette or by synaptic stimulation, and measured GCaMP3 fluorescence with a CMOS camera (Supplementary Figs 10 and 11, and Supplementary Videos 2 and 3) and the same TCSPC-based fibre probes that we used *in vivo* (Supplementary Fig. 12). These experiments indicate that GCaMP3 fluorescence transients represent  $\text{Ca}^{2+}$  responses to suprathreshold synaptic inputs that elicit bursts of action potentials in SPNs.

To examine the relationship between activity in direct- and indirect-pathway SPNs and specific self-paced voluntary actions, we measured GCaMP3 fluorescence while mice performed a lever-pressing operant task. We used a two-lever free-choice task in which mice were rewarded after making ten lever presses, regardless of left- or right-lever pressing. GCaMP3 fluorescence showed transient time-locked increases during the initiation of the session (chamber illumination and lever extension), with a similar temporal profile in both direct and indirect pathways in all animals tested (Fig. 2a–d); this is consistent with previous studies showing that SPNs respond to visual and auditory stimuli<sup>22</sup>. This sensory stimulus-evoked response was not seen in D2-GFP control mice (Fig. 2e).

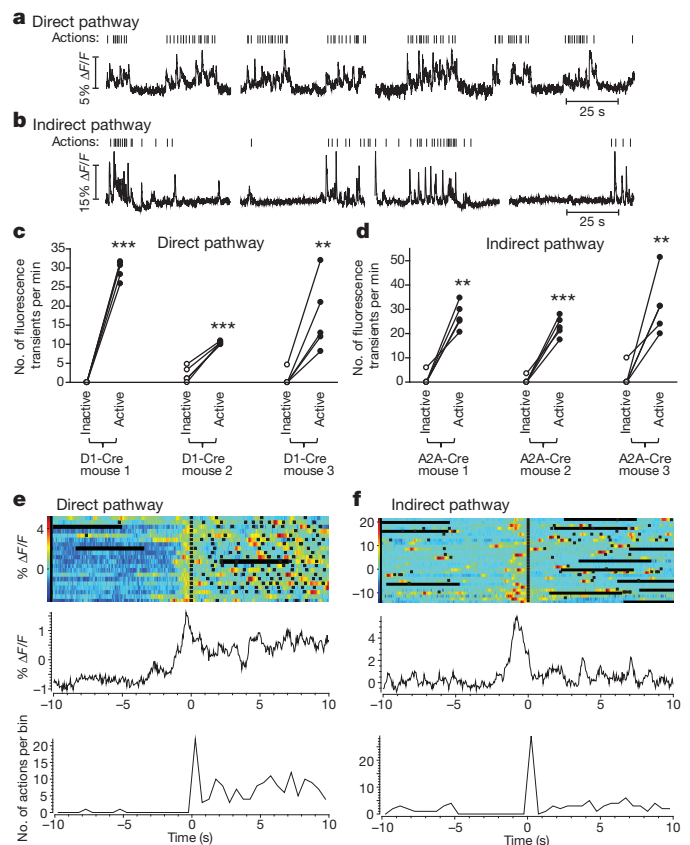
We also observed transient increases in GCaMP3 fluorescence during the performance of different actions (Fig. 2a, b). During a typical operant session, the animals' behaviour was organized into bouts of lever pressing and magazine checking (operationally labelled here as active states) separated by quiet intervals when the mice were not executing operant actions (inactive states) (Fig. 3a, b). If the direct and indirect pathways exert opposing effects on movement, as proposed in the classical model, a higher frequency of transients in direct-pathway SPNs during 'active' states and higher transient frequency in indirect-pathway SPNs during inactive states should be observed. However, both direct- and indirect-pathway SPNs were more active during active states, and quiet during 'inactive' states, as shown by the frequency of GCaMP3 transients during these states (Fig. 3a–d). Although we observed some behaviour-related activation when transients were aligned to individual actions (for example, magazine entries



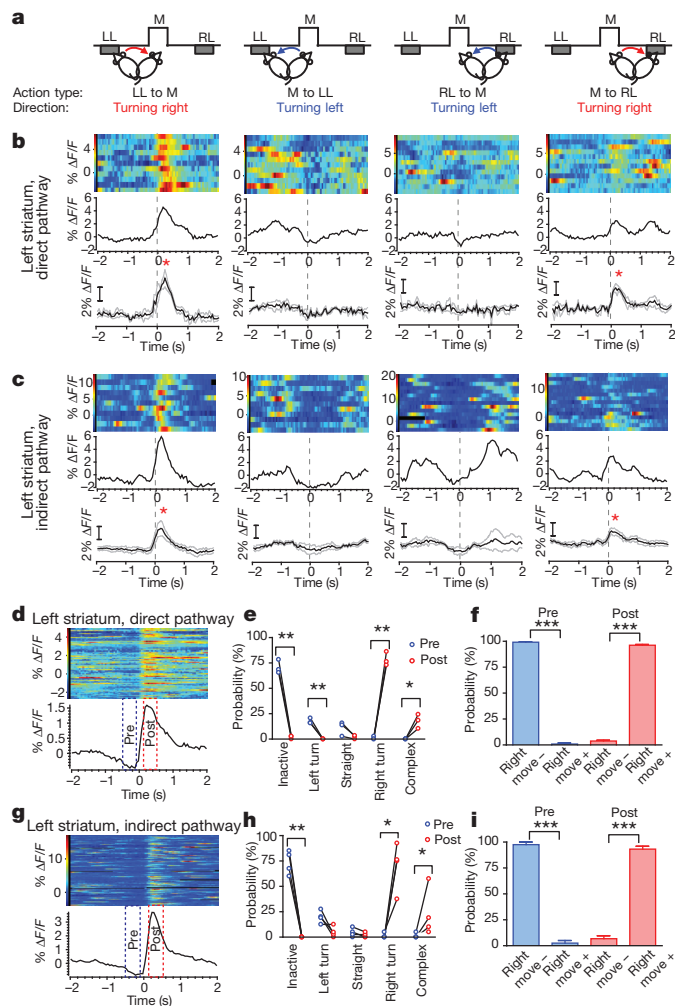
**Figure 2 | Sensory stimuli evoke brief activation in direct- and indirect-pathway SPNs.** a, b, Examples of GCaMP3 fluorescence recorded in the first minute of a lever-pressing session. Vertical ticks indicate time stamps of behavioural events. Arrows indicate session start with complex visual (light on, both levers extended) and auditory (sound of lever extension) stimuli. c–e, Averaged transients in D1-Cre (c), A2A-Cre (d) and D2-GFP mice (e), aligned to session start. \*\*\* $P < 0.001$ , paired  $t$ -test between baseline (averaged between  $-1$  s and session start) and peak value (averaged between 250 ms and 350 ms after session start),  $n = 11$  trials from 4 D1-Cre mice, 11 trials from 4 A2A-Cre mice, and 4 trials from 2 D2-GFP mice. Grey lines indicate  $\pm$  s.e.m.

or lever presses; Supplementary Fig. 14), the strongest phasic activation occurred before the initiation of a set of actions or action sequences (Fig. 3e, f), consistent with previous studies<sup>23</sup>.

To investigate in detail the relationship between these transients and behaviour, we examined fluorescence transients in direct- and indirect-pathway SPNs in relation to the initiation of particular actions. As SPN activity during action generation can depend on the direction and acceleration or speed of the movement, as well as the motivational state of the animal<sup>4,24</sup>, we compared activity during four frequent types of actions in the operant box that were roughly equivalent in acceleration and speed (Supplementary Fig. 13) but had different directions and different motivational goals (initiating a bout of operant presses after magazine entry, or approaching the magazine after a pressing bout, Fig. 4a and Supplementary Video 1). Actions 'left lever to magazine' (LL to M) and 'right lever to magazine' (RL to M) were presumably equal in motivational state but opposite in movement direction. Similarly, action 'magazine to left lever' (M to LL) and 'magazine to right lever' (M to RL) are another pair of actions with similar motivational states but opposite movement directions.



**Figure 3 | GCaMP3 fluorescence transients in both direct- and indirect-pathway SPNs during action initiation in a lever-pressing task.** a, b, Examples of GCaMP3 fluorescence changes in direct- (a) and indirect-pathway (b) SPNs during a two-lever free-choice operant task. Vertical ticks are time stamps for operant actions (lever presses + magazine entries). c, d, Comparison of GCaMP3 transient frequency during 'inactive' states and 'active' states in direct- (c) and indirect-pathway (d) SPNs. Main effect of action states, direct pathway  $F_{1,12} = 212$ ,  $P < 0.001$ ; indirect pathway  $F_{1,12} = 108$ ,  $P < 0.001$ ; post-hoc tests \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ,  $n = 5$  trials for each mouse. e, f, Peri-event time histograms (PETHs) of actions (bottom panel) and corresponding GCaMP3 fluorescence (top panel) of direct-pathway SPNs (e) and indirect-pathway SPNs (f) aligned to the first action in an action sequence. Black squares indicate individual actions. Horizontal black bars are 5-s pauses between 60-s acquisitions.



**Figure 4 | Both direct- and indirect-pathway SPNs show strong activation immediately before and during contraversive movement.** **a**, Illustration of four action types analysed in **b** and **c**: initiation of movement from left lever (LL) to food magazine (M), from M to LL, from right lever (RL) to M, from M to RL. **b, c**, GCaMP3 fluorescence in direct- and indirect-pathway SPNs measured in the left striatum, aligned to initiation of the corresponding actions illustrated in **a**. Top row, multiple trials showing colour-coded GCaMP3 fluorescence and average response from one D1-Cre (**b**) and one A2A-Cre (**c**) mouse; bottom row, averaged responses from three D1-Cre (**b**) and four A2A-Cre mice (**c**). \* $P < 0.05$ , paired-comparison between baseline and GCaMP3 peak. **d–i**, Multiple trials of GCaMP3 fluorescence and average response aligned to the threshold of each detected fluorescence transient from a single mouse (**d, g**). The two boxes indicate the time windows before (Pre) and after (Post) the onset of the transients; these boxes are analysed in panels **e, f, h** and **i**. Probability analyses were carried out of the different types of locomotor activities before and after onset of fluorescence transients in direct- (**e**) and indirect-pathway (**h**) SPNs in left striatum. Interaction between action states, and between Pre and Post; direct pathway,  $F_{4,10} = 233$ ,  $P < 0.001$ ; indirect pathway,  $F_{4,10} = 39.4$ ,  $P < 0.001$ ; post-hoc test, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . Probability analyses (**f, i**) were carried out of two categories of locomotor activity based on whether a rightward movement was made within the defined time window. Right move –, movement status pooled from ‘inactive’, ‘left turn’ and ‘straight’ in **e** and **h**. Right move +, movement status pooled from ‘right turn’ and ‘complex’ in **e** and **h**.  $n = 3$  for D1-Cre,  $n = 4$  for A2A-Cre mice. Error bars represent + s.e.m.

We aligned the GCaMP3 fluorescence signal to the first video frame that showed movement in each action, and observed that both direct- and indirect- pathway SPNs in the left striatum showed strongest activation during rightward movement (LL to M (Fig. 4b, c, first column) and M to RL (Fig. 4b, c, last column)), irrespective of motivational state. This was not observed for leftward movement, regardless of the

motivational states (RL to M and M to LL, (Fig. 4b, c, second and third columns)). However, leftward movements were associated with increased transients in the right striatum (Supplementary Fig. 13). Thus, self-paced movement initiation towards the contralateral side is associated with co-activation of both direct and indirect pathways.

The onset of GCaMP3 fluorescence transients in both pathways preceded movement onset. To test further the relationship between direct- and indirect-pathway SPN activity and the initiation of specific contraversive movement, we aligned the video frames to the onset of the GCaMP3 transients and compared the movement status within a 400-ms window before (–500 ms to –100 ms) and after (100 ms to 500 ms) the onset of the transients (Fig. 4d, g). Transients in both direct- and indirect-pathway SPNs of left striatum reliably preceded movement initiation, and predicted the occurrence of a contraversive movement or a complex movement (but not a straight or ipsilateral movement) within 500 ms of the start of the activity transient (Fig. 4e, f, h, i). Conversely, before a fluorescence transient occurred in both pathways there was a high probability that the animal was inactive (or with less probability, performing an ipsilateral movement) (Fig. 4e, h), indicating further that fluorescence transients in both pathways occurred mostly during movement initiation or during transitions, and predicted the occurrence of specific movements.

In summary, we have developed a new *in vivo* fibre-optic technique for monitoring the activity of specific cell types deep in the brain, and have provided the first definitive evidence that direct- and indirect-pathway striatal neurons are co-activated during movement initiation, and are inactive when the animal is not moving. These data call into question the predominant current model of basal ganglia function, which postulates that there should be more activity in the direct pathway during movement than during immobility, and more activity in the indirect pathway during immobility than during movement. We also observed that neuronal activation in both pathways preceded movement initiation (with a latency appropriate for movement control). Finally, we showed that activity in both pathways predicted the occurrence of specific movements within 500 ms of transient initiation. Although these data do not support a pro-kinetic versus anti-kinetic dichotomy in direct and indirect pathways, they are consistent with other models that propose that the coordinated activation of both pathways is important for action selection, or for the precise timing of basal ganglia output<sup>12,13,25</sup>.

Our data on indirect-pathway SPNs may seem to be at odds with recent studies showing that optical activation of indirect-pathway SPNs decreases locomotion<sup>26</sup>, and that ablating<sup>27</sup> or disrupting the function<sup>28</sup> of indirect-pathway SPNs increases locomotion. However, these optogenetic and lesion findings are not inconsistent with models postulating that co-activation of both pathways are important for action selection, with direct-pathway neurons promoting the wanted motor program and indirect-pathway neurons inhibiting competing motor programs. According to these models, during normal behaviour the coordinated activation of particular groups of direct and indirect-pathway SPNs would promote selection and initiation of a particular movement. In contrast, non-selective global activation of an entire population of indirect-pathway SPNs<sup>26</sup> would probably inhibit most motor programs and not only unwanted ones, leading to bradykinesia. Conversely, inhibiting or ablating most indirect-pathway SPNs<sup>27,28</sup> would abolish suppression of unwanted motor programs and induce hyperkinesia. Opposite results could be expected for global direct-pathway manipulations. Future studies aimed at examining the changes in activity of direct- and indirect-pathway SPNs during basal ganglia disease may shed light on the origin of the motor symptoms associated with basal ganglia dysfunction.

The use of deep brain-inserted fibre optics in freely moving mice enabled us to monitor changes in the fluorescence intensity of a GECI reflecting intracellular calcium transients in specific subgroups of SPNs that control voluntary actions. The spectral and lifetime measurements made using TCSPC provide confirmation of the source of

fluorescence transients, which is not possible with simpler intensity-based fluorometry techniques. The same instrumentation (and biosensor expression strategy) could also be used to monitor changes in fluorescence emission spectra and/or fluorescence lifetime. We envision using this methodology with other fluorophores to examine key downstream changes in protein–protein interactions (using fluorescence-lifetime imaging microscopy and Förster (fluorescence) resonance energy transfer (FLIM-FRET)<sup>29</sup>) and metabolism (using spectral FLIM<sup>30</sup>) that these calcium transients trigger to control behaviour.

## METHODS SUMMARY

Experiments were approved by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) Animal Care and Use Committee, and were in accordance with National Institutes of Health (NIH) guidelines. GCaMP3 was expressed in direct- and indirect-pathway SPNs by microinjection of AAV vectors containing FLEX-GCaMP3 into dorsal striatum of D1-Cre and A2A-Cre mice. Operant training began with a two-lever free-choice fixed-ratio 1 schedule (one reward per press), then progressed to fixed-ratio 5, and finally reached fixed-ratio 10. *In vivo* optical recording was carried out during the fixed-ratio 10 schedule. The fluorescence signal was acquired at 20 Hz by a 16-channel TCSPC system (Becker & Hickl GmbH). The time stamps of behavioural events automatically detected by the operant chamber (Med-Associates) were fed online into a Plexon MAP system. All behavioural tests were also recorded by a video camera at 30 frames per s for detailed behaviour analysis. Data were analysed offline using Neuroexplorer (Plexon), Matlab, Prism and Image J.

**Full Methods** and any associated references are available in the online version of the paper.

Received 13 April; accepted 18 December 2012.

Published online 23 January 2013.

- Albin, R. L., Young, A. B. & Penney, J. B. The functional anatomy of basal ganglia disorders. *Trends Neurosci.* **12**, 366–375 (1989).
- DeLong, M. R. Primate models of movement disorders of basal ganglia origin. *Trends Neurosci.* **13**, 281–285 (1990).
- Mink, J. W. The basal ganglia and involuntary movements: impaired inhibition of competing motor patterns. *Arch. Neurol.* **60**, 1365–1368 (2003).
- Hikosaka, O., Takikawa, Y. & Kawagoe, R. Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiol. Rev.* **80**, 953–978 (2000).
- Alexander, G. E. & Crutcher, M. D. Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci.* **13**, 266–271 (1990).
- Gerfen, C. R. & Surmeier, D. J. Modulation of striatal projection systems by dopamine. *Annu. Rev. Neurosci.* **34**, 441–466 (2011).
- Gerfen, C. R. *et al.* D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science* **250**, 1429–1432 (1990).
- Schiffmann, S. N., Jacobs, O. & Vanderhaeghen, J. J. Striatal restricted adenosine A2 receptor (RDC8) is expressed by enkephalin but not by substance P neurons: an *in situ* hybridization histochemistry study. *J. Neurochem.* **57**, 1062–1067 (1991).
- Schiffmann, S. N., Fisone, G., Moresco, R., Cunha, R. A. & Ferre, S. Adenosine A2A receptors and basal ganglia physiology. *Prog. Neurobiol.* **83**, 277–292 (2007).
- Gong, S. *et al.* Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* **27**, 9817–9823 (2007).
- Nambu, A. Seven problems on the basal ganglia. *Curr. Opin. Neurobiol.* **18**, 595–604 (2008).
- Brown, P. Abnormal oscillatory synchronisation in the motor system leads to impaired movement. *Curr. Opin. Neurobiol.* **17**, 656–664 (2007).
- Chan, C. S., Surmeier, D. J. & Yung, W. H. Striatal information signaling and integration in globus pallidus: timing matters. *Neurosignals* **14**, 281–289 (2005).
- Fuller, D. R., Hull, C. D. & Buchwald, N. A. Intracellular responses of caudate output neurons to orthodromic stimulation. *Brain Res.* **96**, 337–341 (1975).
- Parent, A. *et al.* Organization of the basal ganglia: the importance of axonal collateralization. *Trends Neurosci.* **23**, S20–S27 (2000).
- Lima, S. Q., Hromadka, T., Znamenskiy, P. & Zador, A. M. PINP: a new method of tagging neuronal populations for identification during *in vivo* electrophysiological recording. *PLoS ONE* **4**, e6099 (2009).
- Zariwala, H. A. *et al.* A Cre-dependent GCaMP3 reporter mouse for neuronal imaging *in vivo*. *J. Neurosci.* **32**, 3131–3141 (2012).
- Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* **6**, 875–881 (2009).
- Dombeck, D. A., Harvey, C. D., Tian, L., Looger, L. L. & Tank, D. W. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neurosci.* **13**, 1433–1440 (2010).
- Kerr, J. N. & Denk, W. Imaging *in vivo*: watching the brain in action. *Nature Rev. Neurosci.* **9**, 195–205 (2008).
- Schnütgen, F. *et al.* A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. *Nature Biotechnol.* **21**, 562–565 (2003).
- Hikosaka, O., Sakamoto, M. & Usui, S. Functional properties of monkey caudate neurons. II. Visual and auditory responses. *J. Neurophysiol.* **61**, 799–813 (1989).
- Jin, X. & Costa, R. M. Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature* **466**, 457–462 (2010).
- Samejima, K., Ueda, Y., Doya, K. & Kimura, M. Representation of action-specific reward values in the striatum. *Science* **310**, 1337–1340 (2005).
- Goldberg, J. H. & Fee, M. S. A cortical motor nucleus drives the basal ganglia-recipient thalamus in singing birds. *Nature Neurosci.* **15**, 620–627 (2012).
- Kravitz, A. V. *et al.* Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature* **466**, 622–626 (2010).
- Durieux, P. F. *et al.* D2R striatopallidal neurons inhibit both locomotor and drug reward processes. *Nature Neurosci.* **12**, 393–395 (2009).
- Bateup, H. S. *et al.* Distinct subclasses of medium spiny neurons differentially regulate striatal motor behaviors. *Proc. Natl Acad. Sci. USA* **107**, 14845–14850 (2010).
- Elangovan, M., Day, R. N. & Periasamy, A. Nanosecond fluorescence resonance energy transfer-fluorescence lifetime imaging microscopy to localize the protein interactions in a single living cell. *J. Microsc.* **205**, 3–14 (2002).
- Schweitzer, D. *et al.* Towards metabolic mapping of the human retina. *Microsc. Res. Tech.* **70**, 410–419 (2007).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank C. R. Gerfen for gifts of multiple bacterial artificial chromosome (BAC) transgenic mouse lines; L. L. Looger and the Howard Hughes Medical Institute (HHMI) for permission to use AAV GCaMP3 vectors and GCaMP3 mice; S. R. Ikeda for assistance with Ca<sup>2+</sup> imaging in brain slices; G. Luo for mouse genotyping; C. Thaler for assistance with FLIM curve analysis; B. Mathur and M. Davis for assistance with brain slice electrophysiology and histology; and A. Martin for assistance with AAV vector injection. This work was supported by the Division of Intramural Clinical and Biological Research of the NIAAA, European Research Council STG 243393, an International Early Career Scientist grant from the Howard Hughes Medical Institute to R.M.C., a National Research Foundation of Korea grant (2011-0029485, 2012-0004003) and Smart IT Convergence System Research Center (SIRC-2011-0031866) from the Korean government (MEST) to S.B.J., and by an Ellison Medical Foundation grant (AG-NS-0944-12) to X.J.

**Author Contributions** R.M.C. and S.S.V. conceived the original idea of using the TCSPC technique for optical measurements in freely moving mice. G.C., D.M.L. and R.M.C. designed the experiments. G.C. and S.B.J. set up equipment and optimized procedures for *in vivo* optical recording. G.C. and S.B.J. carried out the *in vivo* experiments and analysed data. G.C. performed *in vitro* experiments and analysed data. X.J. helped with programming and data analysis. M.D.P. performed initial *in vitro* experiments using the TCSPC system and analysed data. G.C., S.S.V., D.M.L. and R.M.C. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.S.V. ([stevevog@mail.nih.gov](mailto:stevevog@mail.nih.gov)), D.M.L. ([lovindav@mail.nih.gov](mailto:lovindav@mail.nih.gov)) or R.M.C. ([ruicosta@fchcampalimad.org](mailto:ruicosta@fchcampalimad.org)).



## METHODS

**Animals.** All animal protocols were approved by the US National Institute on Alcohol Abuse and Alcoholism Animal Care and Use Committee. Experiments were carried out using 3- to 12-month-old male and female mice. D1-Cre, A2A-Cre and D2-GFP bacterial artificial chromosome (BAC) transgenic mice were obtained from the Gene Expression Nervous System Atlas (GENSAT; founder line EY217 for D1-Cre and KG139 for A2A-Cre). Floxed-stop GCaMP3 reporter mice (Ai38) were purchased from the Jackson Laboratory and were crossed with A2A-Cre mice for immuno-staining and *in vitro* slice physiology.

**Viral Expression of GCaMP3 in direct- and indirect-pathway striatal neurons.** To achieve Cre-dependent expression of GCaMP3 in direct- and indirect-pathway SPNs, AAV vectors containing FLEX-GCaMP3 (AAV2/9.hSynap.Flex.GCaMP3.3.SV40, made by Penn Vector Core, lot no. V2178, titre  $1.26 \times 10^{13}$  genome copies per ml) were micro-injected unilaterally into the left or right dorsal striatum by stereotaxic surgery. During the surgery, a small skull window (1 mm  $\times$  1 mm) above the injection site was opened with a dental drill. A 30-gauge injection needle was lowered to the dorsal striatum (tip coordinates from Bregma: anteroposterior +0.5 mm, mediolateral +1.5 mm for left, -1.5 mm for right, dorsoventral -2.5 mm from brain surface). A total volume of 0.8  $\mu$ l of AAV vector per site was injected in 5–10 min. The needle was left in place for a further 10 min before removal. The skull window was then covered with dental acrylic. Animals were allowed to recover for at least 2 weeks before experiments proceeded. We checked the GCaMP3 signals *in vivo* at different time points after virus injection. Although the GCaMP3 spectrum started to appear 2 weeks after injection, it was highly contaminated by the background autofluorescence, indicated by the low ratio of peak GCaMP3 to peak autofluorescence (Supplementary Fig. 4). The peak GCaMP3 to peak autofluorescence ratio grew higher over time and stabilized after 2 months (Supplementary Fig. 4). Thus, for *in vivo* GCaMP3 measurements included in this study, we used animals that received virus injection at least 2 months before the measurements.

**System design for *in vivo* measurement of activity-dependent fluorescence changes in behaving mice.** The system consists of three interconnected modules: the behaviour and control module (Med-Associates), the Plexon MAP system (Plexon) and the TCSPC-based optical measurement system (Becker & Hickl) (Fig. 1a). The behaviour and control module controls the outputs (for example, levers and house light) of the operant chamber and feeds the time stamps of the automatically detected behavioural events to the Plexon system. The Plexon system records all the behavioural-event time stamps and can be used for simultaneous electrophysiological recording when needed. The behaviour and control module also sends out commands to synchronize the TCSPC system and the Plexon system, so that the acquired fluorescence signal can be precisely aligned with the recorded behaviour events (temporal shifts  $\pm$  10 ms). The optical measurement system is a multi-wavelength TCSPC system that simultaneously records the fluorescence spectrum, intensity and lifetime. Within the optical system, a 473-nm pulsed laser is used to excite the brain tissue at 20 MHz through a polarization-maintaining single-mode fibre (although non-polarization-maintaining single-mode fibres are also suitable for use as the excitation fibre, we used polarization-maintaining fibres in our system for applications that are not included in this study). Single-mode fibres were specifically used because the 473-nm light output from these fibres forms a stable, near-perfect Gaussian beam profile, which is required for establishing a stable and reproducible excitation volume at the end of the fibre. At 20 MHz the continuous-wave-equivalent power measured at the free end of the single-mode fibre was approximately 0.1–0.12 mW. Photons emitted from the tissue were collected using a multi-mode fibre and dispersed into spectra by a polychromator. The resulting spectra were projected onto a 16-channel PMT array (PML-16C, Becker & Hickl). The individual photons detected in each PMT channel are recorded by a TCSPC module (SPC-830, Becker & Hickl). The 16-channel PMT array covers a 106-nm wavelength window that can be adjusted anywhere 300 nm and 850 nm using a set screw. The shape and location of the spectrum, and the fluorescence lifetime were used to confirm that detected photons were emitted from GCaMP3. The integrated photon count was used as a measure of intensity. Only animals with high GCaMP3 expression (a peak<sub>GCaMP3</sub> to peak<sub>autofluorescence</sub> ratio higher than 5) and signature lifetime (approximately 2.1 ns) (Fig. 1b–d) were used in behaviour tests and optical measurements.

**Fabrication of the hybrid optical-fibre probe.** A jacketed (outside diameter of 0.9-mm) polarization-maintaining fibre patch cord (core/cladding  $\varnothing$  3.5/125  $\mu$ m, OZ Optics) and a jacketed (outside diameter of 0.9-mm) multi-mode fibre patch cord (core/cladding  $\varnothing$  105/125  $\mu$ m, Thorlabs) were used for fluorescence excitation and detection, respectively. To make the hybrid probe, a jacket (approximately 15 mm long) and acrylic buffer were stripped off at the free ends of the two fibres to expose the cladding. Then fibre ends (approximately 5 mm long) were cleaved off with a fibre cleaver, leaving 10-mm flat-cleaved naked fibre ends.

Under a dissection microscope, the two exposed fibre ends were placed parallel to each other and pressed tightly together. Small adjustments were made to align the tips so that the two fibres stopped at the same plane without staggering. A thin layer of acrylic glue was applied to the fibre ends 5 mm away from the tips. The bound fibre ends then went through a 5-mm long 19- or 20-gauge protective metal tube until the metal tube met with the fibre jackets, leaving 5-mm long naked fibre ends protruding from the other end of the metal tube. More acrylic glue was applied to both ends of the metal tube to ensure a tight bond between the jackets, fibres and the metal tube (Supplementary Fig. 2a, b).

**Effects of  $\text{Ca}^{2+}$  concentration on GCaMP3 fluorescence intensity and lifetime *in vitro*.** Striata from an AAV FLEX-GCaMP3-injected A2A-Cre mouse and a non-injected A2A-Cre mouse (a wild-type control) were dissected and sonicated in two 1.5-ml Eppendorf tubes filled with 1.0 ml patch pipette internal solution (containing in mM: 135 KMeSO<sub>4</sub>, 4 MgCl<sub>2</sub>, 10 HEPES, 0.4 GTP-sodium salt, 4 ATP-Na<sub>2</sub>, 10 phosphocreatine disodium salt, pH 7.2) and 1:100 diluted protease inhibitor cocktail (Sigma-Aldrich), then centrifuged at 13,000 r.p.m. for 10 min. Supernatant (500  $\mu$ l) from each tube was transferred into new 1.5-ml tubes and kept on ice before measurement. During the measurement, a hybrid fibre probe (see description above) was attached to the vertical arm of a mouse stereotaxic (David Kopf Instruments) and lowered into the supernatant (approximately 1 mm below surface). Each sample was measured with increasing exogenous  $\text{Ca}^{2+}$  concentrations (0, 1, 10 and 100  $\mu$ M). The  $\text{Ca}^{2+}$  concentration was increased by adding 1  $\mu$ l of 0.5, 5 and 50 mM CaCl<sub>2</sub> stock solution into the 500- $\mu$ l supernatant. Five measurements were taken for each  $\text{Ca}^{2+}$  concentration. Fluorescence intensity was calculated as the total photon count from eight peak spectral channels (channels 6–13, corresponding approximately to 497–543 nm) integrated over 10 s during each measurement. Fluorescence lifetime constant  $\tau$  was calculated by fitting the fluorescence decay curve with a single exponential decay equation.

**Measuring the approximate excitation and detection depth in the mouse brain tissue using the hybrid fibre probe.** The theoretical excitation and detection volume using two parallel fibres is the overlapping volume of the two acceptance cones formed at the free fibre ends. For the multi-mode detection fibre, the acceptance angle  $\theta_{\text{max}}$  can be calculated from the numerical aperture of the fibre and the refractive index  $n$  of the sample using the equation  $n \cdot \sin \theta_{\text{max}} = \text{numerical aperture}$ . Since the average refractive index of brain tissue is between 1.35 and 1.37 (refs 31, 32), the numerical aperture of the multi-mode fibre used in this study is 0.22, the full acceptance angle  $\theta_{\text{max}}$  of the detection fibre is estimated to be around 19 degrees. As the acceptance angle of the polarization-maintaining excitation fibre cannot be accurately deduced from its numerical aperture number, and because the brain tissue causes strong light scattering, we designed two experiments to estimate empirically the actual excitation and detection depth using fluorescent brain tissue. In the first experiment, we sacrificed a D2-GFP mouse and excised the brain, glued the brain on a plate and sliced it using a vibratome until a coronal section containing the striatum was seen clearly. Then the brain was transferred to a dish containing artificial cerebrospinal fluid (aCSF; see below for formula) under a fluorescence microscope. The hybrid fibre probe was inserted through the cortex and stopped in the dorsal striatum. The striatum was illuminated by the laser beam coming out of the probe and pictures showing GFP fluorescence were taken using a cooled charge-coupled device (CCD) camera (DP71, Olympus, Japan). The approximate volume of fluorophore that can be excited by the polarization-maintaining fibre was estimated using these images (Supplementary Fig. 2c). In the second experiment, we measured the maximum depth of detection of our TCSPC system using a method shown in Supplementary Fig. 2d–f. The rationale for this experiment is that when the fibre probe is moving down vertically through a block of brain tissue with homogeneously distributed fluorophores, the detected photon numbers should be constant until the 'bottom' of the detection cone reaches the bottom surface of the brain block. Proceeding beyond this depth will reduce the total number of fluorophores residing within the detection cone and result in a decrease of the detected photon numbers. The detection depth of the probe was estimated as the distance between two planes: the plane at which the photon counts start to decrease and the plane at which the photon counts are close to zero. Results from these two experiments suggest that in the fluorescent brain tissue, the detection volume had a depth of approximately 500  $\mu$ m.

Thus, a feature of the technique used here is that it reflects the average activity of specific neural populations, and hence, activation of a small number of neurons during particular behaviours may not be detected easily compared to the coordinated activity of large numbers of neurons. Therefore, this technique may be more suitable to test specific predictions of whole-circuit behaviour than to examine the detailed behaviour of small subpopulations of neurons within a particular cell type. This feature would predict that if only a small percentage of indirect-pathway SPNs are active during movement initiation we would not be able to detect them,

and if a large proportion of neurons are active when the animal is not moving we would detect them (and the reverse for the direct-pathway SPNs).

**Two-lever free-choice operant behaviour training.** Behaviour training and optical measurements were carried out in an open-top mouse operant chamber ( $21.6 \times 17.8 \times 12.7$  cm, Med-Associates) housed in a custom-made sound attenuating box. A recessed food magazine flanked by two retractable levers was located close to the bottom of a chamber wall. A 3-W house light was mounted on the wall opposite to the food magazine. Food was restricted throughout the training and mice were maintained at 85% of their original body weight by adjusting the amount of regular laboratory chow given at the end of each day. The training started with a two-lever free-choice fixed-ratio 1 schedule, in which the animals obtained one food reward (20 mg purified pellets from Bio-Serv) after each left- or right-lever press. Each fixed-ratio 1 session lasted 90 min or until mice received 30 rewards. Once the animals were able to obtain 30 rewards within 90 min, the training progressed to a two-lever free-choice fixed-ratio 5 schedule, in which a total of 5 lever presses, regardless of left or right, were needed for each reward delivery. The training moved on to two-lever free-choice fixed-ratio 10 schedule after the animals could obtain 30 rewards in 90 min in ratio 5 sessions. The animals were then trained in fixed-ratio 10 sessions until they could obtain 30 rewards in 30 min. On average, it took 5 to 7 days of training, 1 to 2 sessions per day, for animals to be ready for optical measurement. Some of the animals trained in the two-lever free-choice task developed strong bias towards one of the two levers during the training, making it difficult to obtain within-subject comparison of left-lever and right-lever pressing behaviours. In these cases, we further trained the animals in pairs of single-lever FR10 sessions during which either the left or right lever was presented. For each trained animal, the optical measurements were carried out in two consecutive sessions with either left or right lever presented. To minimize the impact of satiety levels between sessions, only 10 rewards were obtained by the animal in each session.

**Procedures to lower the fibre probe into the dorsal striatum for *in vivo* optical measurement.** On the day of optical measurement, brief surgery was carried out to lower the optical fibre probe into the dorsal striatum while the animal was under isoflurane anaesthesia. The dental acrylic covering the previously drilled skull window was carefully removed with a dental drill until the window was fully exposed. The hybrid optical-fibre probe was slowly lowered into the cortex along the lateral edge of the skull window towards the dorsal striatum at approximately 200  $\mu$ m per step until the GCaMP3 spectrum was detected. The probe was then lowered at 50  $\mu$ m per step until the fluorescence intensity reached a plateau. The final tip location was anteroposterior +0.5 mm, mediolateral +2.0 mm (left), -2.0 mm (right) from Bregma, approximately 2 mm below the brain surface. The probe was then fixed in place with a generous quantity of dental acrylic to ensure a sturdy bond between the probe and the skull. The optical-fibre-tethered animal was transferred to the operant chamber and recovered for 2 h under supervision.

***In vivo* optical measurement in mice performing operant tasks.** The recording session began with a session-start signal, house lights being turned on, and both left and right levers being extended into the chamber. The spectrum data were recorded continuously at 20 Hz for 1 min then paused for 5 s for system buffering. The 1-min acquisition and 5-s buffering were repeated 30 times during the operant session or until the system received a session-end signal from the behaviour recording and control module (if the animal obtained 30 rewards within 30 min). The percentage  $\Delta F/F$  was calculated by  $100 \times (F - F_{\text{mean}})/F_{\text{mean}}$ , where  $F_{\text{mean}}$  was the mean fluorescence intensity throughout an entire 1-min acquisition fragment. Fluorescence intensity was calculated as the total photon count from 8 peak spectral channels integrated over 50 ms. Spectral channels 6–13 (approximately 497–543 nm) and 5–12 (approximately 490–537 nm) were selected for GCaMP3 and GFP, respectively, owing to the slight red shift of the GCaMP3 emission spectrum relative to GFP (Fig. 1c). The detection threshold for a fluorescence transient (Figs 3c, d and 4d–i) was defined as  $\mu + 3\sigma$ , where  $\mu$  and  $\sigma$  were the mean and the standard deviation of the 500-ms fluorescence baseline period preceding the interrogated data point. The 'inactive states' (Fig. 3c, d), were defined as a 15-s period with no detected operant actions. An action sequence (Fig. 3e, f) was defined as a series of action events (magazine entries and lever presses) with inter-event interval of less than 5 s. In the probability analysis (Fig. 4e, h), probability was calculated from 30 to 40 fluorescence transients per animal. 'Left turn' and 'right turn' were defined as unidirectional body movements with rotation of the long axis of more than 30 degrees, accompanied by head turning towards the same direction. 'Straight' was defined as body movement with less than a 30-degree change in the direction of the long axis and without head turn. 'Complex' was defined as rapid movements with abrupt shift in left and right direction.

**Effects of isoflurane anaesthesia on fluorescence signals in the striatum.** Optical-fibre-tethered animals were first placed in a novel cage with fresh bedding

and baseline activity was recorded for 5 min. They were then transferred into an anaesthesia-induction chamber continuously infused with an isoflurane and O<sub>2</sub> mixture. The O<sub>2</sub> flow rate was set at 1.5–2 l min<sup>-1</sup>. The isoflurane vaporizer (E-Z Anaesthesia, E-Z Systems Corp) setting was initially set at 5, then adjusted to between 5 and 1 to maintain deep anaesthesia in the animals. After recording for 5 min in the chamber, the animals were placed back in the novel cage and recorded for a further 20 min during recovery.

**Effects of isoflurane anaesthesia on striatal local-field potentials.** The isoflurane anaesthesia procedures were the same as in the previous paragraph. Microwire electrode arrays used for recording striatal local-field potentials and single units consisted of two rows of eight polyamide-coated tungsten microwires with platinum-plated tips (35  $\mu$ m diameter, CD Neural Technologies). The two rows were separated by 200  $\mu$ m. The inter-electrode spacing within each row was 150  $\mu$ m. The procedures for implanting microwire electrode arrays and recording in behaving animals have been described previously<sup>33</sup>. During the surgery, the centre of the array was placed at 0.5 mm rostral to the bregma, 1.5 mm left to the bregma. The array was lowered through a 1 mm  $\times$  2 mm skull window and inserted 2.5 mm below the brain surface, then fixed in place with dental acrylic. Animals were allowed to recover for at least 2 weeks before experiments started. Local-field potentials were recorded using the MAP system (Plexon) and analysed using Neuroexplorer (Plexon).

**Simultaneous GCaMP3 Ca<sup>2+</sup> imaging and cell-attached firing recording in acutely cut brain slices.** Coronal slices (250- $\mu$ m thick) containing the striatum from 1- to 2-month-old A2A-Cre; floxed-stop GCaMP3 double-mutant mice were cut using a vibratome (Leica VT1200) in cold sucrose cutting solution containing (in mM): 194 sucrose, 30 NaCl, 4.5 KCl, 1 MgCl<sub>2</sub>, 26 NaHCO<sub>3</sub>, 1.2 NaH<sub>2</sub>PO<sub>4</sub>, 10 glucose and incubated at 35 °C for 30 min in aCSF containing the following (in mM): 124 NaCl, 4.5 KCl, 1.2 NaH<sub>2</sub>PO<sub>4</sub>, 1 MgCl<sub>2</sub>, 2 CaCl<sub>2</sub>, 10 glucose, 26 NaHCO<sub>3</sub>, saturated with 95% O<sub>2</sub> and 5% CO<sub>2</sub> (final pH, 7.4; osmolality, 310 mOsm kg<sup>-1</sup>). The slices were then transferred to room temperature (approximately 24 °C) before recording. All recordings were carried out at room temperature. A Neo sCMOS camera (Andor Technology) installed on a SliceScope (Scientifica) upright microscope was used for imaging. The slices were visualized with a water-immersion objective ( $\times 40$ , numerical aperture of 0.8) using either differential interference contrast (DIC) optics or fluorescence excitation with a mercury lamp and GFP filters. Andor SOLIS software was used to capture and analyse the images. To find healthy GCaMP3-expressing cells that were responsive to synaptic activation, a brief train (5 pulses at 100 Hz) of stimulating pulses (0.2 ms, 500  $\mu$ A) was applied through a concentric bipolar electrode (inner pole diameter of 25  $\mu$ m, CBARC75 from FHC) before attempting patch-clamp recordings. Many cells that exhibited no detectable fluorescence under resting conditions showed fluorescence transients during and after stimulation. The location of these cells was marked and patch-clamp recordings were established on them under DIC optics. All electrophysiological recordings and current injections were carried out in tight-seal ( $>2$ -G $\Omega$ ) cell-attached mode<sup>34</sup> using a Multiclamp 700A amplifier (Molecular Devices). Data were acquired using Clampex 8.0 (Molecular Devices), filtered at 5 kHz and digitized at 10 kHz. The patch-pipette solution contained (in mM): 126 K-Gluconate, 4 KCl, 10 HEPES, 4 ATP-Mg, 0.3 GTP-Na, 10 Phosphocreatine (pH 7.2, 290 Osm kg<sup>-1</sup>). To measure synaptic stimulation-evoked GCaMP3 fluorescence transients using the TCSPC-based fibre-optics system, the hybrid fibre probe (see descriptions above) was positioned on the stage with the tip touching the surface of the slice at an approximately 45-degree angle. The stimulating electrode was placed approximately 50–100  $\mu$ m away from the tip of the fibre probe. The time-resolved spectra were recorded at 20 Hz.

**Double immunostaining.** Brains of AAV FLEX-GCaMP3-injected A2A-Cre mice and A2A-Cre;FLEX-GCaMP3 double-mutant reporter mice were first fixed by transcardial perfusion of 4% PFA, then stored in 4% PFA at 4 °C overnight. Coronal slices (30  $\mu$ m) were cut using a microtome (VIBRATOME 1000 Plus) and blocked with 4% horse serum before being incubated in primary antibodies at 4 °C overnight. The primary antibodies used were: chicken anti-GFP (1:1000, Abchem) for detecting GCaMP3, rabbit anti-ChAT (1:1000, Millipore) for detecting ChAT; rabbit anti-Met-ENK (1:1000, Millipore) for detecting Met-ENK; and rabbit anti-TH (1:1000, Millipore) for detecting tyrosine hydroxylase. After washing out excessive primary antibodies, the slices were incubated in the secondary antibodies for 2 h at room temperature. The secondary antibodies used were Alexa488 conjugated goat anti-chicken (1:1000, Invitrogen) and Alexa568 conjugated goat anti-rabbit (1:1000, Invitrogen). Cells were visualized using a Zeiss Axiocam MR monochrome camera installed on Axiovert 200 fluorescent microscope with a  $\times 10$  (numerical aperture 0.3) or a  $\times 20$  (numerical aperture 0.8) objective. Images were acquired using Axiovision 4.8 software (Zeiss).

**Imaging acute brain slices for verification of GCaMP3 expression and location of the fibre-probe tip.** Sagittal slices (300- $\mu$ m thick) containing the striatum from

AAV-injected D1-Cre and A2A-Cre mice were cut serially using a vibratome (Leica VT1200) in cold aCSF containing the following (in mM): 126 NaCl, 2.5 KCl, 1.2 NaH<sub>2</sub>PO<sub>4</sub>, 1.2 MgCl<sub>2</sub>, 2.4 CaCl<sub>2</sub>, 11 glucose, 21.4 NaHCO<sub>3</sub>, saturated with 95% O<sub>2</sub> and 5% CO<sub>2</sub> (final pH 7.4, osmolarity 300 mOsm kg<sup>-1</sup>). The slices were then transferred to a dish and examined under a fluorescence microscope (Olympus MVX10, with ×1 MVX Plan Apochromat Lens, numerical aperture 0.25). Two pictures were taken for each slice. GFP filter was used to show virally expressed GCaMP3. A brightfield channel was used to show the general brain structures. Images were merged offline in Image J.

**Statistical analysis.** The statistical analysis was carried out using GraphPad Prism 4 (GraphPad Software). All averaged data were reported as mean ± s.e.m. Comparisons using Fisher's partial least-squares difference (PLSD) or paired *t*-test

were carried out after verifying a main effect or interaction using two-way analysis of variance (ANOVA), with repeated measures when appropriate.

31. Lue, N. *et al.* Live cell refractometry using Hilbert phase microscopy and confocal reflectance microscopy. *J. Phys. Chem. A* **113**, 13327–13330 (2009).
32. Binding, J. *et al.* Brain refractive index measured in vivo with high-NA defocus-corrected full-field OCT and consequences for two-photon microscopy. *Opt. Express* **19**, 4833–4847 (2011).
33. Costa, R. M., Cohen, D. & Nicolelis, M. A. Differential corticostriatal plasticity during fast and slow motor skill learning in mice. *Curr. Biol.* **14**, 1124–1134 (2004).
34. Perkins, K. L. Cell-attached voltage-clamp and current-clamp recording and stimulation techniques in brain slices. *J. Neurosci. Methods* **154**, 1–18 (2006).



# A direct and melanopsin-dependent fetal light response regulates mouse eye development

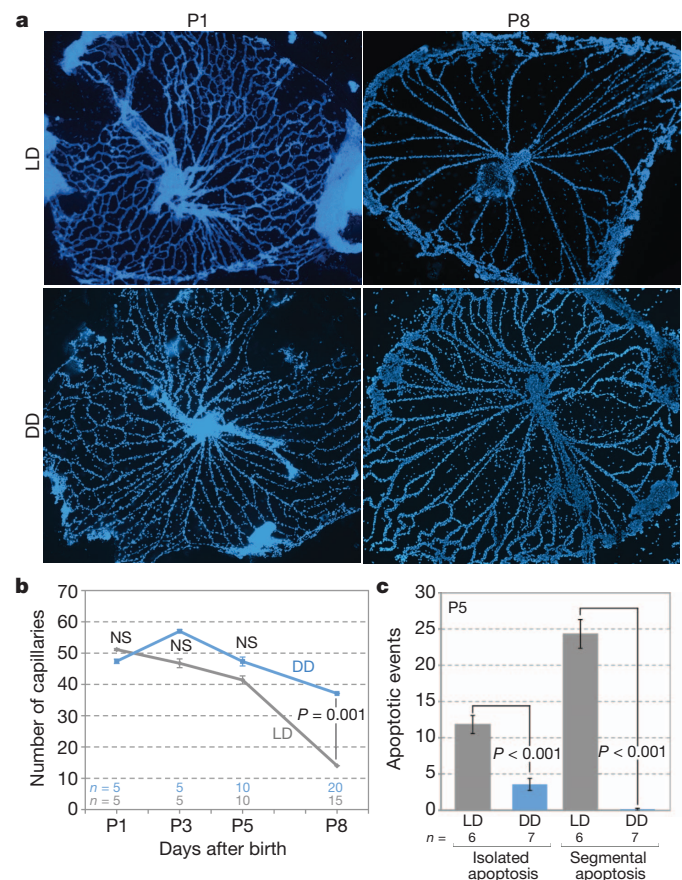
Sujata Rao<sup>1</sup>, Christina Chun<sup>2</sup>, Jieqing Fan<sup>1</sup>, J. Matthew Kofron<sup>3</sup>, Michael B. Yang<sup>1,4</sup>, Rashmi S. Hegde<sup>3</sup>, Napoleone Ferrara<sup>5</sup>, David R. Copenhagen<sup>2</sup> & Richard A. Lang<sup>1,3,4</sup>

**Vascular patterning is critical for organ function. In the eye, there is simultaneous regression of embryonic hyaloid vasculature<sup>1</sup> (important to clear the optical path) and formation of the retinal vasculature<sup>2</sup> (important for the high metabolic demands of retinal neurons). These events occur postnatally in the mouse. Here we have identified a light-response pathway that regulates both processes. We show that when mice are mutated in the gene (*Opn4*) for the atypical opsin melanopsin<sup>3–5</sup>, or are dark-reared from late gestation, the hyaloid vessels are persistent at 8 days post-partum and the retinal vasculature overgrows. We provide evidence that these vascular anomalies are explained by a light-response pathway that suppresses retinal neuron number, limits hypoxia and, as a consequence, holds local expression of vascular endothelial growth factor (VEGFA) in check. We also show that the light response for this pathway occurs in late gestation at about embryonic day 16 and requires the photopigment in the fetus and not the mother. Measurements show that visceral cavity photon flux is probably sufficient to activate melanopsin-expressing retinal ganglion cells in the mouse fetus. These data thus show that light—the stimulus for function of the mature eye—is also critical in preparing the eye for vision by regulating retinal neuron number and initiating a series of events that ultimately pattern the ocular blood vessels.**

Prompted by the recognition that newborn mice are light-responsive<sup>6</sup> and show light-dependent neuronal connectivity changes<sup>7</sup>, we proposed the existence of a pathway in which light responsiveness in the early retina might trigger hyaloid regression and thus clearance of the optic axis. To test this, we placed pregnant dams in the dark at late gestation (embryonic (E) day 16–17): pups raised in the dark until 8 days post partum (P8) showed persistent hyaloid vessels (Fig. 1a) and this was confirmed by quantification over a P1–P8 time course (Fig. 1b). Assessment of hyaloid vessel numbers at P15 showed that by this stage they had regressed. This indicated that dark-rearing resulted in a regression delay. Quantification of apoptosis at P5 showed that regardless of whether we quantified the isolated events that predominate early in hyaloid regression or the segmental pattern of apoptosis that follows<sup>8</sup>, there was a reduction (Fig. 1c) similar quantitatively to previously characterized hyaloid persistence mutants<sup>9</sup>. These data suggest that a light-response pathway promotes hyaloid regression.

Hyaloid vessel regression and superficial layer retinal angiogenesis occur at the same time in the mouse<sup>1,10</sup>, and this indicated that dark-rearing might affect both processes. Retinal angiogenesis in mice begins at the day of birth with the extension of vessel precursors from the head of the optic nerve. A superficial layer of vasculature within the retinal ganglion cell (RGC) layer extends to the retinal periphery by P7 (Supplementary Fig. 1a–c). Starting at about P8, angiogenic sprouts extend vertically downwards into the deeper layers of the retina (Supplementary Fig. 1d) and ultimately form the deep vasculature at the outer edge of the inner nuclear layer and the intermediate plexus

within the inner plexiform layer<sup>2</sup>. In mice dark-reared from E16 to E17, the superficial vascular plexus (Supplementary Fig. 1e) showed an increase in density regardless of whether the region was simple plexus (Supplementary Fig. 1f) or at a vein (Supplementary Fig. 1g). Depth-coded P8 image stacks (Supplementary Fig. 1h) showed that there were many more descending vessels (Supplementary Fig. 1h, yellow, blue) than in the wild type (Supplementary Fig. 1h, yellow) and many of these were abnormally located. These changes were confirmed by



**Figure 1 | Hyaloid regression is regulated by light.** **a**, Hyaloid vessel preparations at the indicated postnatal (P) days from pups reared under normal light conditions (LD) or under constant darkness (DD) from E16–17. Original magnification,  $\times 50$ . **b**, As in **a** but a quantification of vessel number from P1 to P8.  $P$  values obtained by analysis of variance (ANOVA). **c**, P5 apoptotic index in hyaloid vascular cells (isolated apoptosis) or vessels undergoing a segmental pattern of apoptosis.  $P$  values obtained by Student's  $t$ -test. Sample size ( $n$ ) as labelled. NS, not significant. Error bars are s.e.m.

<sup>1</sup>The Visual Systems Group, Abrahamson Pediatric Eye Institute, Division of Pediatric Ophthalmology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. <sup>2</sup>Departments of Ophthalmology and Physiology, University of California, San Francisco, San Francisco, California 94158, USA. <sup>3</sup>The Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. <sup>4</sup>Department of Ophthalmology, University of Cincinnati, College of Medicine, Cincinnati, Ohio 45229, USA. <sup>5</sup>Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA.

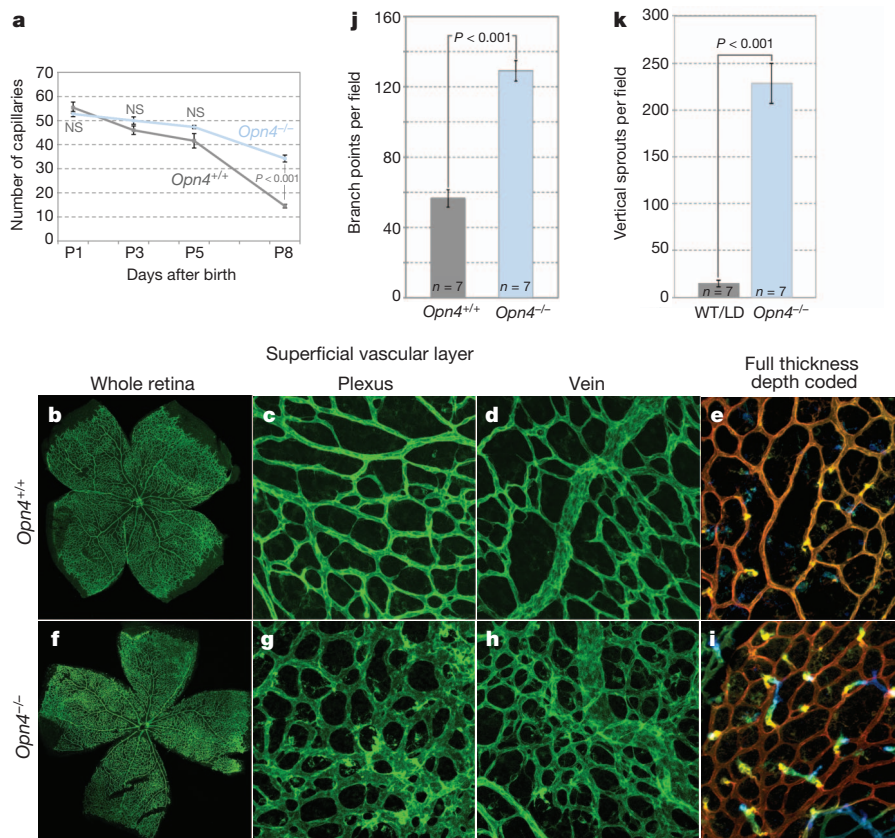
quantification (Supplementary Fig. 1i, j). Thus, the retinal vasculature is a second vascular structure in the eye where normal development is disrupted by the absence of light.

Melanopsin is expressed from an early stage of both mouse and human gestation and unlike photoreceptor opsins, is known to function in the mouse eye before P10 (ref. 11). Melanopsin-expressing intrinsically photosensitive retinal ganglion cells (ipRGCs) are a subset of RGCs that function in circadian entrainment and the pupillary reflex<sup>12</sup>. ipRGCs are located in the superficial layers of the retina adjacent to both the retinal and hyaloid vasculatures (Supplementary Fig. 2a–c). This location, the pre-photoreceptor functions of melanopsin<sup>6</sup> and the vascular anomalies present in mice that are missing RGCs<sup>13</sup>, suggested that it was a good candidate to mediate light-dependent vascular development in the eye. To test this possibility, we assessed hyaloid vessel regression and retinal vascular development in mice mutated in *Opn4*, the melanopsin-encoding gene<sup>4,5</sup>. *Opn4*<sup>−/−</sup> mice showed normal hyaloid vessel numbers at P1 but persistence at P8 (Fig. 2a and Supplementary Fig. 2d). Examination of P15 eyes showed that hyaloid regression was complete in the *Opn4*<sup>−/−</sup> mice, indicating that, as with dark-reared mice, hyaloid persistence was not long term. *Opn4*<sup>−/−</sup> mice also showed a retinal vascular overgrowth phenotype that qualitatively and quantitatively (Fig. 2b–k) mimicked the changes resulting from dark rearing. To determine whether changes in retinal vascular density endured, we performed a quantitative assessment at P15, P25 and P180 (Supplementary Fig. 3) that showed that elevated vascular density was regionally sustained until at least P180. More generally, the vascular phenotype of the *Opn4*<sup>−/−</sup> mice phenocopies that observed in dark-reared mice. This provides an independent means of implicating a light-response pathway in

vascular development of the eye and identifies melanopsin as the opsin required.

VEGFA is a potent signal for vascular endothelial cell survival<sup>14</sup> that is required for retinal angiogenesis<sup>15</sup> and is also present in the vitreous of the rodent<sup>16</sup> and human<sup>17</sup> eye where the hyaloid vessels reside. We proposed that light-dependent vascular development might be explained by modulation of VEGFA. Consistent with this, homozygous and heterozygous deletion of *Vegfa*<sup>1</sup> with the *Chx10-cre* retinal driver<sup>18</sup> gave, respectively, either a hyaloid development failure (Supplementary Fig. 4a) or diminished hyaloid regression (Fig. 3a and Supplementary Fig. 4b). An immunoblot for vitreous VEGFA over the P1–P8 time course revealed that in control mice, VEGFA164 levels were reduced at P5 but rose again by P8 (Fig. 3b). When three different time courses of VEGFA immunoblots were quantified, the P5 VEGFA signal was about fivefold reduced compared with P1 (Fig. 3b). A low level of VEGFA at P5 is consistent with the idea that it is a key regulator of hyaloid regression because P5 is the time when there are peak levels of vascular endothelial cell apoptosis<sup>9</sup>.

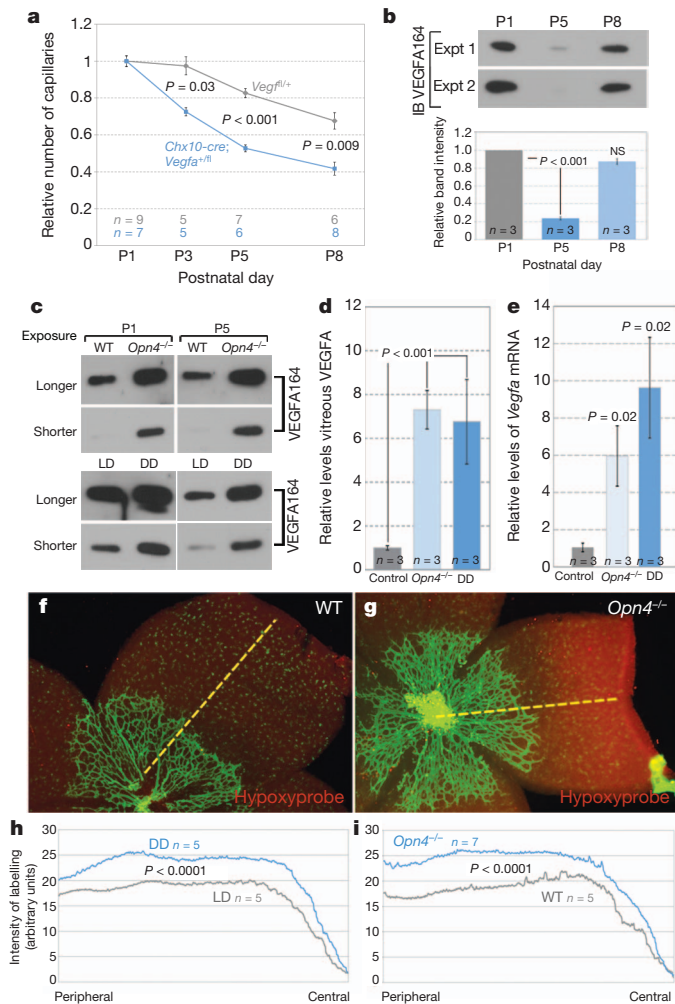
Using dark rearing and the *Opn4*<sup>−/−</sup> mice, we determined whether actual or functional darkness resulted in a modulation of vitreous VEGFA. In four independent experiments we consistently observed that vitreous VEGFA levels were increased regardless of how light responsiveness was compromised (Fig. 3c). Furthermore, an enzyme-linked immunosorbent assay (ELISA)-based assessment of VEGFA in the P5 vitreous showed that whether pups were dark-reared or mutated in *Opn4*, the levels of VEGFA were about sevenfold higher than in the control (Fig. 3d). A sevenfold increase in VEGFA in the vitreous of dark-reared and *Opn4*<sup>−/−</sup> mice was reflected in similar fold increases in the level of retinal *Vegfa* messenger RNA as indicated by



**Figure 2 | Hyaloid regression and retinal angiogenesis are regulated by melanopsin.** **a**, Quantification of hyaloid vessels in *Opn4*<sup>+/+</sup> and *Opn4*<sup>−/−</sup> mice over a P1 to P8 time course.  $P$  values obtained by ANOVA. **b–i**, Low (×100; **b**, **f**) and high (×200; **c–e**, **g–i**) magnification images of isolectin-labelled P8 retina from wild-type (**b–e**) and *Opn4*<sup>−/−</sup> (**f–i**) pups raised in

normal lighting. **e**, **i**, Depth-coded z stack images for wild type (**e**) and *Opn4*<sup>−/−</sup> (**i**) indicate the appearance of vertical angiogenic sprouts. **j**, **k**, Graphs show quantification of branch points (**j**) and vertical sprouts (**k**) in animals of the indicated genotypes. WT, wild type.  $P$  values obtained by Student's  $t$ -test. Errors bars are s.e.m. Sample sizes ( $n$ ) as labelled.

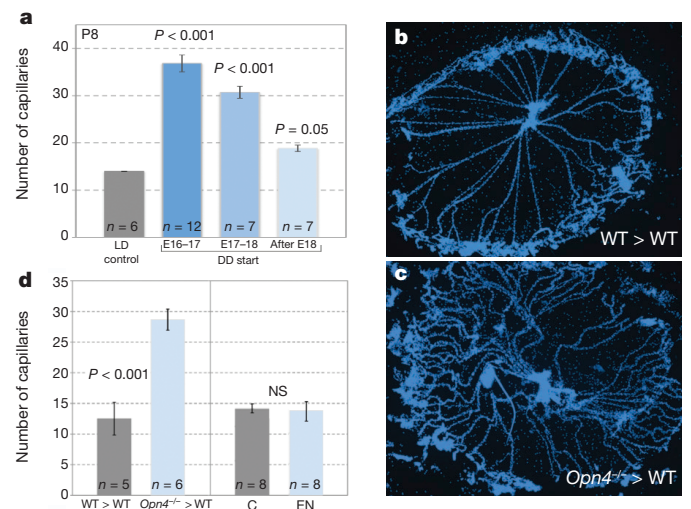




**Figure 3 | Light and melanopsin-dependent regulation of VEGFA expression and hypoxia in the retina.** **a**, Hyaloid vessel number from P1 to P8 in mice of labelled genotypes. *P* values obtained by ANOVA. **b**, Vitreous VEGFA immunoblot (IB) for wild-type mice at P1, P5 and P8 with quantification histogram. **c**, Immunoblot for P1 or P5 vitreous VEGFA in wild-type and *Opn4*<sup>-/-</sup> mice or in mice reared in LD or DD light conditions as labelled. **d**, ELISA quantification of VEGFA levels in the P5 vitreous of control/LD mouse pups (grey bar) from *Opn4*<sup>-/-</sup> mice (pale blue bar) or those raised in constant darkness from E16–17 (DD, blue bar). **e**, qPCR detection of *Vegfa* mRNA in P5 retina from control/LD (grey bar), *Opn4*<sup>-/-</sup> mice (light blue bar) and dark-reared mice (DD, dark blue bar). *P* values in **b**, **d**, **e** were obtained by Student's *t*-test. Sample sizes (*n*) as labelled. Error bars are s.e.m. **f**, **g**, Labelling of flat-mount P5 retinas from wild-type (**f**) and *Opn4*<sup>-/-</sup> (**g**) mice for blood vessels (isolectin, green) and for hypoxia (red). Retinal myeloid cells label faintly with isolectin. Original magnification,  $\times 100$ . **h**, **i**, Quantification of the relative levels of hypoxyprobe labelling in the retinas of LD and DD mice (**c**) and wild-type versus *Opn4*<sup>-/-</sup> (**d**) retinas.

quantitative polymerase chain reaction (qPCR) (Fig. 3e). Flow sorting/qPCR further showed that Thy1.1<sup>+</sup> RGCs and Thy1.1<sup>-</sup> Vc1.1<sup>+</sup> amacrine/horizontal cells exhibited an increase in *Vegfa* mRNA, although Thy1.1<sup>-</sup> Vc1.1<sup>-</sup> PDGFR<sup>+</sup> astrocytes did not (Supplementary Fig. 5a). Given the VEGFA dependence of hyaloid vessels (Supplementary Fig. 4) and of retinal angiogenesis<sup>15</sup>, elevated retinal VEGFA expression is an explanation for the vascular anomalies observed in dark-reared and *Opn4*<sup>-/-</sup> mice.

Quantification of BRN3B<sup>+</sup> RGCs and calretinin<sup>+</sup> amacrine cells in P5 *Opn4*<sup>-/-</sup> mice revealed modest increases in the numbers of both cell types (Supplementary Fig. 5d). It has been shown that retinal angiogenesis in the mouse is driven by a hypoxia-response pathway that upregulates VEGFA expression<sup>19,20</sup>. Because increased cell



**Figure 4 | Gestational light controls vascular development in the eye.**

**a**, Quantification of hyaloid vessels in mice raised in normal lighting (LD, grey bar) and those dark-reared from E16–17 (dark-blue bar), E17–18 (medium-blue bar), or after E18 (light-blue bar). **b**, **c**, P8 hyaloid vessel preparations from a wild-type embryo transferred into a wild-type pseudopregnant female (WT>WT) and an *Opn4*<sup>-/-</sup> embryo transferred into a wild-type pseudopregnant female (*Opn4*<sup>-/-</sup>>WT). Original magnification,  $\times 50$ . **d**, Left panel: quantification of hyaloid vessels in P8 WT>WT (*n* = 5) and *Opn4*<sup>-/-</sup>>WT (*n* = 6) pups. Right panel: quantification of hyaloid vessels in normal control pups at P8 (C; *n* = 8) and P8 pups (*n* = 8) born to an enucleated female (EN). Sample sizes (*n*) as labelled. *P* values in **a** by ANOVA; *P* values in **d** were obtained by Student's *t*-test. Error bars are s.e.m.

number can increase oxygen demand, we tested whether the *Opn4* mutation and dark rearing resulted in retinal hypoxia (Fig. 3f, g). Quantification of labelling with hypoxyprobe at P5 (Fig. 3h, i) revealed that increased signal was a consequence of both *Opn4* mutation and dark rearing. This is consistent with the hypothesis that elevated VEGFA expression in the retina is a consequence of increased oxygen demand due to higher numbers of retinal neurons (Supplementary Fig. 7).

In assessing the role of light responses in vascular development of the eye, our presumption had been that birth was probably a triggering event because vascular patterning events occur after birth and because light levels to the eye of the newborn would increase. To test this, we dark-reared pups from different points of late gestation (E16–17, E17–E18 or after E18) and assessed hyaloid persistence. We observed a dose response where the hyaloid vessels were progressively more persistent with an earlier dark-rearing start (Fig. 4a). In particular, if dark rearing was started after E18 (the day of birth is usually E19) there was almost no effect (Fig. 4a, light blue bar). These data indicate the surprising outcome that the critical light-response period stimulating hyaloid regression is gestational, at or before E16–17. This further raised the possibility that this developmental pathway required a direct fetal light response. To test this assertion directly, we transferred embryos from an *Opn4*<sup>+/-</sup>  $\times$  *Opn4*<sup>+/-</sup> cross into a pseudopregnant wild-type female and assessed hyaloid regression. Control, wild-type pups in the transferred litter showed normal hyaloid regression (Fig. 4d, b, grey bar), whereas the *Opn4*<sup>-/-</sup> mice showed persistence at P8 (Fig. 4d, c, blue bar). In addition, to test the reciprocal possibility—that the light response of the mother might influence vascular development of the fetal eye—we enucleated female mice, and allowed them to gestate and raise litters under normal lighting conditions. This did not produce hyaloid persistence (Fig. 4d). Combined, these experiments show that melanopsin in the fetus, not the mother, is critical to regulate vascular development of the eye.

Measurements of the light level in the visceral cavity of adult mice living under mouse room fluorescent lights (Supplementary Fig. 6)



revealed flux densities of  $1.4 \times 10^{13}$  (for BALB/c) and  $1.1 \times 10^{12}$  (for C57BL/6) photons  $\text{cm}^{-2} \text{s}^{-1}$ . Published response thresholds for rodent ipRGCs range upwards from approximately  $1.2 \times 10^{10}$  photons  $\text{cm}^{-2} \text{s}^{-1}$  (refs 21–23). Furthermore, it has been shown that ipRGCs have the ability to respond continuously to light stimulation, via melanopsin, for up to 10 h (ref. 23). Although it has been suggested that ipRGCs in newborn mouse pups are less sensitive than in adults<sup>22</sup>, the reduced sensitivity is about 1.5 log quanta and so the visceral light level in a pigmented animal of  $1.1 \times 10^{12}$  photons  $\text{cm}^{-2} \text{s}^{-1}$  may still be above the threshold. These data are consistent with the hypothesis that the mouse fetus can respond directly to light via melanopsin.

These experimental studies identify light as a trigger for hyaloid vessel regression and for suppression of promiscuous angiogenesis in the retina. The observation that dark rearing from late gestation or *Opn4* mutation produces essentially identical perturbations of vascular development provides corroborating evidence for involvement of a melanopsin-dependent, light-response pathway (Supplementary Fig. 7). Our data also indicate that the origin of the hyaloid persistence and de-regulated retinal angiogenesis is increased levels of VEGFA originating in retinal neurons. These findings are surprising because, with the exception of neuronal connectivity, it has not been shown previously that light can trigger changes in developmental programs. Our data indicate that the primary light-dependent change is an increase in the number of retinal neurons and that the vascular changes occur in response to increased oxygen demand (Supplementary Fig. 7) considerably later in developmental time. This pathway is an interesting example of one where events unfold slowly over the course of nearly 2 weeks. It will be interesting to determine whether this pathway might influence susceptibility to retinopathy of prematurity<sup>24</sup>, the retinal vasculopathy of pre-term infants in which promiscuous angiogenesis can cause blindness.

## METHODS SUMMARY

We prepared and labelled the retinal vasculature and hyaloid vessels as reported previously<sup>9,25</sup>. VEGFA was detected in the vitreous of *Opn4* mutant and dark-reared mice using standard immunoblotting and ELISA (R&D) techniques. Retinal neurons were identified and enumerated using standard techniques of immunofluorescence labelling. The level of retinal hypoxia in *Opn4* mutant and dark-reared mice was assessed using detection of injected pimonidazole hydrochloride (Hypoxyprobe). All animal experiments were performed in accordance with IACUC-approved guidelines and regulations.

**Full Methods** and any associated references are available in the online version of the paper.

Received 25 April 2011; accepted 5 December 2012.

Published online 16 January 2013.

1. Ito, M. & Yoshioka, M. Regression of the hyaloid vessels and pupillary membrane of the mouse. *Anat. Embryol. (Berl.)* **200**, 403–411 (1999).
2. Fruttiger, M. Development of the retinal vasculature. *Angiogenesis* **10**, 77–88 (2007).
3. Provencio, I., Jiang, G., De Grip, W. J., Hayes, W. P. & Rollag, M. D. Melanopsin: an opsin in melanophores, brain, and eye. *Proc. Natl Acad. Sci. USA* **95**, 340–345 (1998).
4. Hattar, S., Liao, H. W., Takao, M., Berson, D. M. & Yau, K. W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002).

5. Panda, S. *et al.* Melanopsin (*Opn4*) requirement for normal light-induced circadian phase shifting. *Science* **298**, 2213–2216 (2002).
6. Johnson, J. *et al.* Melanopsin-dependent light avoidance in neonatal mice. *Proc. Natl Acad. Sci. USA* **107**, 17374–17378 (2010).
7. Tian, N. & Copenhagen, D. R. Visual stimulation is required for refinement of ON and OFF pathways in postnatal retina. *Neuron* **39**, 85–96 (2003).
8. Diez-Roux, G. & Lang, R. A. Macrophages induce apoptosis in normal cells *in vivo*. *Development* **124**, 3633–3638 (1997).
9. Lobov, I. B. *et al.* WNT7b mediates macrophage-induced programmed cell death in patterning of the vasculature. *Nature* **437**, 417–421 (2005).
10. Saint-Geniez, M. & D'Amore, P. A. Development and pathology of the hyaloid, choroidal and retinal vasculature. *Int. J. Dev. Biol.* **48**, 1045–1058 (2004).
11. Tarrtlin, E. E. *et al.* Expression of opsin genes early in ocular development of humans and mice. *Exp. Eye Res.* **76**, 393–396 (2003).
12. Do, M. T. & Yau, K. W. Intrinsically photosensitive retinal ganglion cells. *Physiol. Rev.* **90**, 1547–1581 (2010).
13. Edwards, M. M. *et al.* The deletion of Math5 disrupts retinal blood vessel and glial development in mice. *Exp. Eye Res.* **96**, 147–156 (2012).
14. Ferrara, N., Gerber, H. P. & LeCouter, J. The biology of VEGF and its receptors. *Nature Med.* **9**, 669–676 (2003).
15. Haigh, J. J. *et al.* Cortical and retinal defects caused by dosage-dependent reductions in VEGF-A paracrine signaling. *Dev. Biol.* **262**, 225–241 (2003).
16. Meeson, A. P., Argilla, M., Ko, K., Witte, L. & Lang, R. A. VEGF deprivation-induced apoptosis is a component of programmed capillary regression. *Development* **126**, 1407–1415 (1999).
17. Adamis, A. P. *et al.* Increased vascular endothelial growth factor levels in the vitreous of eyes with proliferative diabetic retinopathy. *Am. J. Ophthalmol.* **118**, 445–450 (1994).
18. Rowan, S. & Cepko, C. L. Genetic analysis of the homeodomain transcription factor Chx10 in the retina using a novel multifunctional BAC transgenic mouse reporter. *Dev. Biol.* **271**, 388–402 (2004).
19. Kurihara, T. *et al.* von Hippel-Lindau protein regulates transition from the fetal to the adult circulatory system in retina. *Development* **137**, 1563–1571 (2010).
20. Lange, C. *et al.* Retina-specific activation of a sustained hypoxia-like response leads to severe retinal degeneration and loss of vision. *Neurobiol. Dis.* **41**, 119–130 (2010).
21. Sekaran, S. *et al.* Melanopsin-dependent photoreception provides earliest light detection in the mammalian retina. *Curr. Biol.* **15**, 1099–1107 (2005).
22. Tu, D. C. *et al.* Physiologic diversity and development of intrinsically photosensitive retinal ganglion cells. *Neuron* **48**, 987–999 (2005).
23. Wong, K. Y. A retinal ganglion cell that can signal irradiance continuously for 10 hours. *J. Neurosci.* **32**, 11478–11485 (2012).
24. Connor, K. M. *et al.* Quantification of oxygen-induced retinopathy in the mouse: a model of vessel loss, vessel regrowth and pathological angiogenesis. *Nature Protocols* **4**, 1565–1573 (2009).
25. Stefater, J. A. III *et al.* Regulation of angiogenesis by a non-canonical Wnt–Ft1 pathway in myeloid cells. *Nature* **474**, 511–515 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank D. Bredl, P. Speeg and M. Sarangdhar for technical assistance, A. Delwig and N. Brown for advice. We acknowledge the assistance of the Research Flow Cytometry Core in the Division of Rheumatology at CCHMC, supported in part by NIH AR-47363. This work was supported by the NIH (R.A.L., D.R.C., J.M.K. and R.S.H.) with additional funding from the Abrahamson Pediatric Eye Institute of CCHMC, That Man May See at UCSF, Research to Prevent Blindness (D.R.C.) and March of Dimes (D.R.C.).

**Author Contributions** R.A.L. and D.R.C. provided project leadership. R.A.L., D.R.C. and S.R. wrote the manuscript. R.A.L., D.R.C. and R.S.H. supervised experimental work. N.F. developed critical reagents and M.B.Y. provided important clinical insights. S.R., C.C., J.F., J.M.K. and D.R.C. performed experimentation and analysis.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.A.L. ([richard.lang@cchmc.org](mailto:richard.lang@cchmc.org)) or D.R.C. ([cope@phy.ucsf.edu](mailto:cope@phy.ucsf.edu)).

## METHODS

**Mice.** Genotyping of *Vegfa*<sup>fl</sup> (ref. 26), *Chx10-cre* (ref. 18), *Opn4<sup>cre</sup>* (ref. 27), *Ai4* (ref. 28) and *Opn4<sup>-/-</sup>* (ref. 5) was performed as described. All animal experimentation was carried out using protocols approved by the Institutional Animal Care and Use Committee at Cincinnati Children's Hospital Medical Center and at the University of California San Francisco.

**Hyaloid and retinal labelling and quantification.** Hyaloid vessels were collected and stained with Hoechst as well as for TdT-mediated dUTP nick end labelling (TUNEL) as described<sup>9</sup>. Retinal flat-mounts were prepared and labelled with isolectin<sup>29</sup> or for melanopsin<sup>4</sup>. Hyaloid vessel quantification has been described previously<sup>4</sup>. Retinal vessel density was quantified by counting vessel junctions using ImageJ for many  $\times 200$  microscope fields. Depth-coded three-dimensional image reconstructions were generated using a Zeiss Apotome-equipped microscope in conjunction with Axiovision software. Antibodies for labelling of retinal flat-mounts included anti-Brn3b (Abcam), anti-calretinin (Millipore) and anti-melanopsin (ATS).

**Assessment of hypoxia.** P5 mouse pups were injected with 60 mg kg<sup>-1</sup> (about 180  $\mu$ g per pup) pimonidazole hydrochloride (Hypoxyprom), and after 45 min were killed and retinas collected. The rabbit primary antibody to pimonidazole hydrochloride was then used in conjunction with an anti-rabbit Alexa594 secondary to label retinal tissue. We quantified labelling by generating intensity values along line intervals extending from the retinal centre to the periphery. Pixel intensity values from 20 to 25 line intervals per retina and 5 to 6 retinas per genotype were averaged. Significance values were calculated using the MatLab ANOVA test.

**Isolation and analysis of vitreous.** Vitreous was collected from dark-reared pups in a room using red illumination. Eyes from P1 and P5 pups were washed twice in sterile ice cold PBS. Excess PBS was blotted using a kimwipe, a small slit was made through the retina and vitreous collected. ELISA was performed on the vitreous using the Vegfa Quantikine kits (R&D) that include recombinant protein standards. Immunoblots were probed with a unique carboxy-terminal antibody for VEGFA<sup>30</sup> from Santa Cruz. Quantification was performed using ImageJ.

**Cell sorting.** For flow sorting using markers for retinal neurons<sup>31</sup>, retina was dissociated as described<sup>25</sup> except 16 mg ml<sup>-1</sup> of liberase CI (Roche) and 20  $\mu$ g ml<sup>-1</sup> of DNase I (Sigma) was used. Cells were then labelled on ice for 30 min with goat PDGFR- $\alpha$ , washed with PBS and labelled with PerCP-conjugated anti-CD90 (clone OX-7), FITC-conjugated anti-CD57 (clone VC1.1), Alexa fluor 350 and 7-AAD. Cells were sorted with FACSaria II running DiVa software.

**RNA isolation and qPCR.** RNA was isolated using RNeasy (Qiagen). qPCR was performed with QuantiTect SYBR green (Qiagen) using amplification of actin for

normalization. In analysing qPCR data, the *P* values refer to a comparison of the  $\Delta\Delta$ CT values. Primers were as follows: *Vegfa* 5'-GACAGAACAAAGCCAGA-3', 5'-CACCGCCTTGGCTTGTCAC-3'.

**Light measurements.** To estimate the radiant flux density impinging a mouse housed in a cage in the UCSF animal care facilities, we determined the spectral distribution of fluorescent lights,  $S(\lambda_{\text{relative}})$ , illuminating the room and the absolute power (watts cm<sup>-2</sup> s<sup>-1</sup>) of these lights. Both measurements were done at floor level.  $S(\lambda_{\text{relative}})$  was measured (Photo Research, PR670 spectroradiometer) as light reflected from a spectrally flat reflective surface (Spectralon Target, Labsphere). The radiant power was measured using a calibrated radiometric detector (UDT Instruments; model S471).  $S(\lambda_{\text{relative}})$  was converted to  $S(\lambda_{\text{absolute}})$  by scaling the area under  $S(\lambda_{\text{relative}})$  to match the radiant power and then converting these values to photons cm<sup>-2</sup> s<sup>-1</sup> for each wavelength. The melanopsin spectral absorbance curve was then convolved with  $S(\lambda_{\text{absolute}})$ . The area under this curve was used as a measure of the radiant flux density capable of stimulating the melanopsin pigment ( $\lambda_{\text{max}} = 479$  nm). We calculate that the equivalent of at least  $5 \times 10^{13}$  photons cm<sup>-2</sup> s<sup>-1</sup> was available to stimulate melanopsin. Similar measurements and calculations for sunlight (12:00, 20 December 2011) revealed a radiant flux density of  $2.6 \times 10^{16}$  photons cm<sup>-2</sup> s<sup>-1</sup>.

To estimate attenuation of light that could stimulate melanopsin in mouse fetuses *in utero*, we directed a blue LED (Philips Lumileds Lighting Company; model: Luxeon III star, LXHL-LB3C, peak wavelength = 470 nm) that was positioned 1 inch from the skin to a miniature silicon photodiode light detector placed inside the abdominal cavity. Light penetrated both the skin and sub-dermal muscle layer. The measurements were done on live, anaesthetized adult mice (intramuscular (IM) injections of ketamine/xylazine).

26. Lichtenberger, B. M. *et al.* Autocrine VEGF signaling synergizes with EGFR in tumor cells to promote epithelial cancer development. *Cell* **140**, 268–279 (2010).
27. Ecker, J. L. *et al.* Melanopsin-expressing retinal ganglion-cell photoreceptors: cellular diversity and role in pattern vision. *Neuron* **67**, 49–60 (2010).
28. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neurosci.* **13**, 133–140 (2010).
29. Gerhardt, H. *et al.* VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J. Cell Biol.* **161**, 1163–1177 (2003).
30. Stockmann, C. *et al.* Deletion of vascular endothelial growth factor in myeloid cells accelerates tumorigenesis. *Nature* **456**, 814–818 (2008).
31. Goldberg, J. L., Klassen, M. P., Hua, Y. & Barres, B. A. Amacrine-signaled loss of intrinsic axon growth ability by retinal ganglion cells. *Science* **296**, 1860–1864 (2002).

# In vitro expansion of single *Lgr5*<sup>+</sup> liver stem cells induced by Wnt-driven regeneration

Meritxell Huch<sup>1\*</sup>, Craig Dorrell<sup>2\*</sup>, Sylvia F. Boj<sup>1</sup>, Johan H. van Es<sup>1</sup>, Vivian S. W. Li<sup>1</sup>, Marc van de Wetering<sup>1</sup>, Toshiro Sato<sup>1†</sup>, Karien Hamer<sup>1</sup>, Nobuo Sasaki<sup>1</sup>, Milton J. Finegold<sup>3</sup>, Annelise Haft<sup>2</sup>, Robert G. Vries<sup>1</sup>, Markus Grompe<sup>2</sup> & Hans Clevers<sup>1</sup>

The Wnt target gene *Lgr5* (leucine-rich-repeat-containing G-protein-coupled receptor 5) marks actively dividing stem cells in Wnt-driven, self-renewing tissues such as small intestine and colon<sup>1</sup>, stomach<sup>2</sup> and hair follicles<sup>3</sup>. A three-dimensional culture system allows long-term clonal expansion of single *Lgr5*<sup>+</sup> stem cells into transplantable organoids (budding cysts) that retain many characteristics of the original epithelial architecture<sup>2,4,5</sup>. A crucial component of the culture medium is the Wnt agonist RSPO1<sup>6</sup>, the recently discovered ligand of LGR5<sup>7,8</sup>. Here we show that *Lgr5-lacZ* is not expressed in healthy adult liver, however, small *Lgr5-LacZ*<sup>+</sup> cells appear near bile ducts upon damage, coinciding with robust activation of Wnt signalling. As shown by mouse lineage tracing using a new *Lgr5-IRES-creERT2* knock-in allele, damage-induced *Lgr5*<sup>+</sup> cells generate hepatocytes and bile ducts *in vivo*. Single *Lgr5*<sup>+</sup> cells from damaged mouse liver can be clonally expanded as organoids in Rspo1-based culture medium over several months. Such clonal organoids can be induced to differentiate *in vitro* and to generate functional hepatocytes upon transplantation into *Fah*<sup>-/-</sup> mice. These findings indicate that previous observations concerning *Lgr5*<sup>+</sup> stem cells in actively self-renewing tissues can also be extended to damage-induced stem cells in a tissue with a low rate of spontaneous proliferation.

Quiescent liver stem cells are believed to reside in biliary ducts<sup>9</sup>. Sox9- and Foxl1-based lineage tracing have proven the existence of such cells<sup>10–13</sup>. In the adult liver, the Wnt pathway is exclusively active in hepatocytes that surround central veins (perivenous hepatocytes)<sup>14</sup>. In bile ducts, Wnt signalling becomes active following liver injury<sup>15</sup>. Consistent with these findings, we detected activity of the generic Wnt reporter *Axin2-lacZ* (ref. 16) only in perivenous hepatocytes, and this was upregulated upon induction of liver injury by carbon tetrachloride (CCl<sub>4</sub>) injection<sup>17</sup> (Supplementary Fig. 1a, b). Maximal expression occurred between day 3 to 6 after damage (Supplementary Fig. 1c). By microarray analysis, we noted induction of *Wnt6* (>twofold), of several Rspodins (three- to sixfold) and of many Wnt target genes previously characterized in intestinal crypt cells<sup>8</sup>, including *Lgr5* (≥twofold). Notably, perivenous hepatocyte Wnt target genes (*Gul*, *Slc1a2*, *Rhbg* and *Cyp1a2*)<sup>14</sup> were downregulated, indicating that Wnt activation occurred outside perivenous hepatocytes (Supplementary Fig. 1d and Supplementary Table 1).

In untreated *Lgr5-lacZ* knock-in mice<sup>1</sup>, *Lgr5-lacZ* expression was essentially undetectable (Fig. 1a). Upon CCl<sub>4</sub> treatment, clear reporter activity (peaking at day 5 to 6) occurred in groups of small cells near the bile ducts clearly distinguishable from neighbouring hepatocytes (Fig. 1b and Supplementary Fig. 2a–c). These *Lgr5*<sup>+</sup> cells expressed Sox9, a relatively broad ductal progenitor marker<sup>10,12,13</sup>, but did not express mature hepatocyte or stellate cell markers (Supplementary Fig. 2d–f). The gene expression profile of CCl<sub>4</sub>-induced *Lgr5*<sup>+</sup> cells correlated closely with biliary duct cells but not hepatocytes

(Supplementary Fig. 2g). Closer comparison with the biliary duct profile revealed that multiple Wnt target genes and multiple intestinal stem cell genes<sup>18</sup> were enriched in liver *Lgr5*<sup>+</sup> cells (Supplementary Tables 2 and 3).

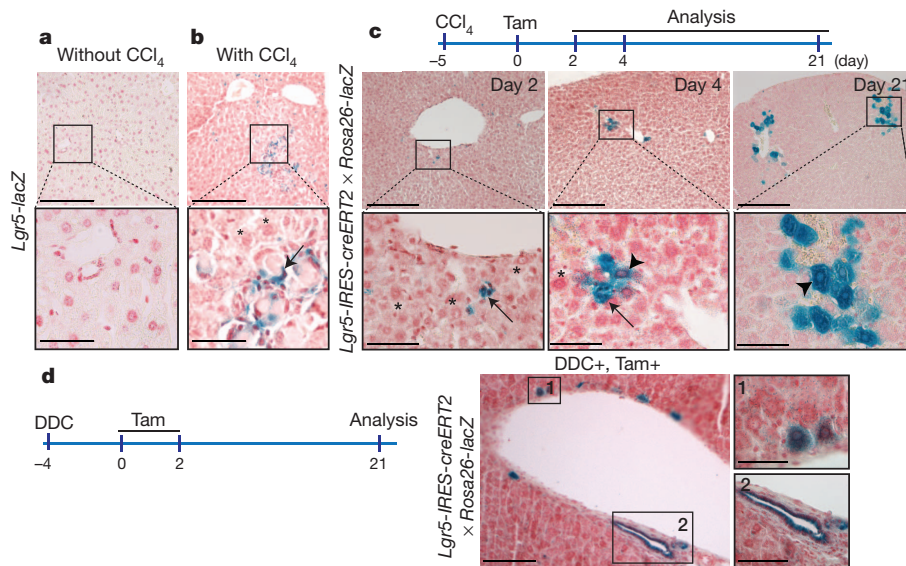
We then aimed to visualize the progeny of liver *Lgr5* cells by lineage tracing. The *Lgr5-EGFP-IRES-creERT2* allele<sup>1</sup> is permanently silenced in liver. Therefore, we generated a new *Lgr5* allele by inserting the *IRES-creERT2* cassette into the 3' untranslated region (UTR) of the *Lgr5* locus rather than in the first exon (Supplementary Fig. 3a), and we crossed these mice with the *Rosa26-lacZ* reporter<sup>19</sup>. After a single tamoxifen injection, lineage tracing events were readily detected in the intestine, validating this allele (Supplementary Fig. 3b). Adult offspring were treated with CCl<sub>4</sub> and 5 days later Cre activity was activated by tamoxifen. Two days after tamoxifen induction, groups of small, proliferative LacZ<sup>+</sup> cells became visible that evolved into fully mature hepatocytes from two days later onwards (Fig. 1c). Because CCl<sub>4</sub> induces central vein damage, we also tested two 'oval cell response' models: MCDE (methionine choline-deficient diet supplemented with ethionine)<sup>20</sup> and DDC (3,5-diethoxycarbonyl-1,4-dihydrocollidine)<sup>21</sup>. In both models, lineage tracing from *Lgr5* cells was readily detected in hepatocytes and biliary ducts (Fig. 1d and Supplementary Fig. 3d–f). In the absence of liver damage, no tracing events were detected in the livers of mice with the same genotype (Supplementary Fig. 3c). Similar tracing data have been reported for *Foxl1* (refs 11, 13).

Given the expression of the Wnt-dependent *Lgr5* stem cell marker, we reasoned that adult liver progenitors could possibly be expanded from the ductal compartment under our previously defined organoid culture conditions<sup>2,4</sup>. Previously established liver culture methods typically yield cell populations that undergo senescence over time<sup>10,13,22–24</sup> unless the cells are transformed. To establish liver progenitor cultures, biliary duct fragments were embedded in Matrigel containing the 'generic' organoid culture factors epidermal growth factor (EGF) and Rspo1<sup>4</sup>, to which Fgf10, HGF and nicotinamide (expansion medium, EM) were added. Virtually all fragments formed cysts that grew into much larger liver organoids (Supplementary Fig. 4a, b) expressing *Lgr5* and ductal markers (Supplementary Fig. 4c). Without EGF, Rspo1 or nicotinamide, the cultures deteriorated within one to two passages (Supplementary Fig. 4d). Cultures have been maintained more than 12 months by weekly passaging at a 1:8 ratio. We then initiated single-cell (clonal) cultures from *Lgr5-LacZ*<sup>+</sup> cells, FACS sorted from *Lgr5-lacZ* mice previously induced with a single dose of CCl<sub>4</sub> (Fig. 2a, b). Sorted cells cultured in our defined expansion medium conditions rapidly divided and formed cyst-like structures that were maintained for more than 8 months by weekly passaging at a 1:8 ratio (Fig. 2c and Supplementary Fig. 5e). Karyotypic analysis of both clonal and bulk cultures, revealed that most cells (~85%) harboured normal chromosome numbers, even at 8 months (Supplementary

<sup>1</sup>Hubrecht Institute for Developmental Biology and Stem Cell Research, Uppsalalaan 8, 3584CT Utrecht & University Medical Centre Utrecht, Netherlands. <sup>2</sup>Oregon Stem Cell Center, Papé Family Pediatric Research Institute, Oregon Health and Science University, Portland, Oregon 97239, USA. <sup>3</sup>Department of Pathology, Texas Children's Hospital, Houston, Texas 77030 USA. <sup>†</sup>Present address: Department of Gastroenterology, School of Medicine, Keio University, 35 Shinanomachi, Shinjuku-ku, Tokyo, 160-8582, Japan.

\*These authors contributed equally to this work.





**Figure 1 | Liver damage induces *Lgr5*<sup>+</sup> bipotential liver progenitors.**

**a, b**, *Lgr5-lacZ* mice were injected i.p. with corn oil ( $n = 6$ ) (**a**) or with a single dose of  $\text{CCl}_4$  in corn oil ( $1 \text{ ml kg}^{-1}$ ) ( $n = 6$ ) (**b**). Six days later, mice were killed and livers collected and processed for  $\beta$ -galactosidase staining. **a**, Undamaged liver does not express *Lgr5-lacZ*. **b**, Following  $\text{CCl}_4$  treatment, strong *Lgr5-lacZ* expression was detected in small cells near ducts. Compare *Lgr5-LacZ*<sup>+</sup> cells (arrow) with neighbouring hepatocytes (asterisks). Scale bars, 200  $\mu\text{m}$  (top) and 30  $\mu\text{m}$  (bottom). **c**, *Lgr5-IRES-creERT2* mice were crossed with *Rosa26-lacZ* Cre reporter mice. Offspring received a single i.p. injection of  $\text{CCl}_4$  and 5 days later CreERT2 activity was induced with tamoxifen (3 mg per mouse) as

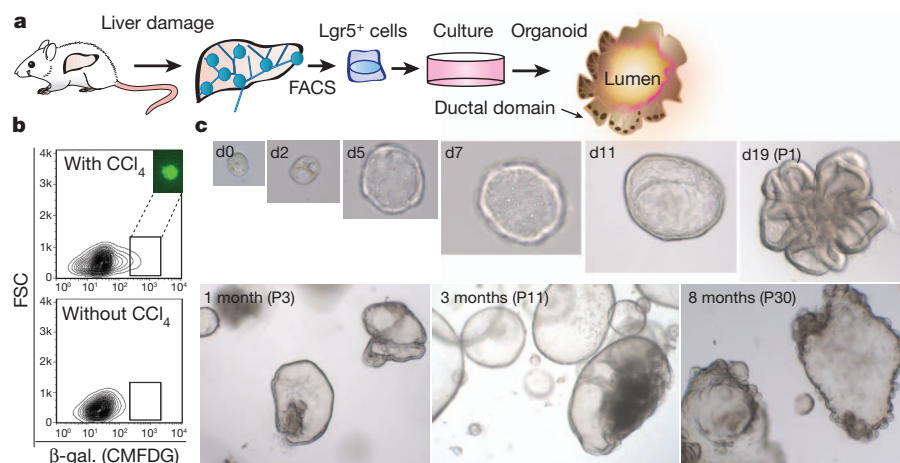
indicated in the scheme. Representative pictures showing lineage tracing from *Lgr5* cells upon  $\text{CCl}_4$  damage ( $n = 8$ ). Compare differences in size between *LacZ*<sup>+</sup> cells (arrows) and hepatocytes (asterisks and arrowheads) at days 2 and 4 after tamoxifen induction. Arrowheads indicate *LacZ*<sup>+</sup> hepatocytes on days 4 and 21. Magnifications on each day are the same. Scale bars, 500  $\mu\text{m}$  (top) and 50  $\mu\text{m}$  (bottom). **d**, *Lgr5-IRES-creERT2*  $\times$  *Rosa26-lacZ* offspring were fed with DDC ( $n = 3$ ), and 4 and 6 days later CreERT2 activity was induced with tamoxifen as indicated in the scheme. Representative pictures showing positive hepatocytes (**d**, inset 1) and positive ductal cells (**d**, inset 2). Tam, tamoxifen. Scale bars, 200  $\mu\text{m}$  (**d**), 25  $\mu\text{m}$  (**d**, inset 1) and 50  $\mu\text{m}$  (**d**, inset 2).

Fig. 4e), consistent with the  $\sim 25\%$  level of aneuploidy in young adult mouse liver<sup>25</sup>. Importantly, secondary cultures from *Lgr5-lacZ*<sup>+</sup> cells could also be established and expanded for more than 4 months in culture (Supplementary Fig. 5a–e).

To assess the lineage potential of *Lgr5* cells, we performed gene expression profiling of clonal organoids. Microarray analysis revealed that clonal organoid expression profiles resembled adult liver. *Lgr5* and progenitor markers such as *Sox9*, *Cd44* and *Prom1* (ref. 10) were highly upregulated. The clonal organoids expressed multiple hepatocyte-lineage markers as well as bile duct markers, revealing that single *Lgr5* cells are bipotential (Supplementary Fig. 6a–f). Markers of mature hepatocytes were only weakly expressed or absent (Supplementary Fig. 7a, EM column).

Marker analysis suggested that the culture conditions were biased towards induction of a biliary cell fate. To induce hepatocyte maturation

*in vitro*, we defined a differentiation medium (DM). Inhibition of Notch and TGF- $\beta$  signalling, both implicated in biliary cell fate determination *in vivo*<sup>26,27</sup>, induced the expression of  $\sim 200$  genes. These included *Tbx3*, *Pparg* and *Bmp2*, genes essential for liver maturation<sup>27–29</sup>, as well as mature hepatocyte markers such as *Cyp3a11*, *Fah*, *G6pc* and *Alb* (Supplementary Fig. 7a, b). We also observed induction of a set of genes involved in cholesterol and lipid metabolism, as well as genes encoding p450 cytochromes (Supplementary Fig. 7c, d). Accordingly, the progenitor profile was shut down, as indicated by downregulation of *Lgr5* (Supplementary Fig. 7a, DM column). Immunofluorescent staining revealed the expression of Hnf4- $\alpha$  and albumin, as well as the basolateral membrane protein Mrp4 (also known as Abcc4) and the tight junction protein ZO-1 (also known as Tjp1) (Fig. 3a–d). Up to 33% of the cells were positive for the hepatocyte marker OC2-2F8 and displayed high granularity by

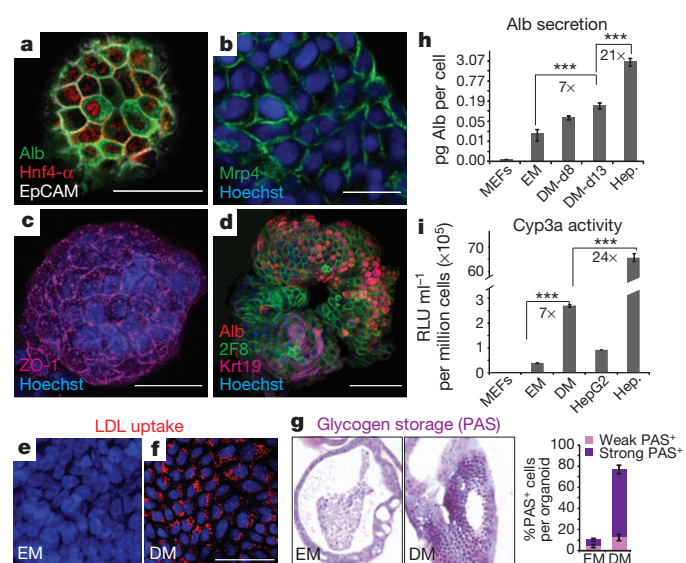


**Figure 2 | In vitro expansion of single *Lgr5* cells from adult liver tissue.**

**a–c**, *Lgr5-lacZ* mice were injected i.p. with corn oil or  $\text{CCl}_4$  as in Fig. 1. Six days later, liver tissue was dissociated to single cells, loaded with the fluorescent  $\beta$ -galactosidase (CMFDG) substrate and analysed by fluorescent-activated cell sorting (FACS). Sorted isolated *Lgr5-LacZ*<sup>+</sup> cells were cultured at a ratio of one single *Lgr5-LacZ*<sup>+</sup> cell per well (clonal) as described in the Methods. **a**, Scheme representing the protocol used. **b**, Representative FACS plot of dissociated single cells from  $\text{CCl}_4$ -treated (with  $\text{CCl}_4$ ) and non-treated (without  $\text{CCl}_4$ ) livers. Cells were gated following sequential selection by cell-size (forward scatter (FSC) versus side scatter (SSC)) and propidium iodide (PI) exclusion. Viable CMFDG<sup>+</sup>PI<sup>−</sup> cells were selected and sorted. Representative sorted cell is shown. **c**, Serial differential interference contrast (DIC) images showing the outgrowth of a single *Lgr5-LacZ*<sup>+</sup> cell. Original magnifications were  $\times 40$  (days 0–5),  $\times 20$  (day 7–11),  $\times 10$  (day 19) and  $\times 4$  (1 month onwards). P, passage.

flow cytometry analysis, a feature of mature hepatocytes (Fig. 3d and Supplementary Fig. 7e). Binucleation, a hallmark of hepatocyte maturation, was also detected (Supplementary Fig. 7f). Of note, the ductal phenotype was not fully abolished, as patches of Krt19-positive cells remained present (Fig. 3d). The differentiated organoids were subjected to several tests for hepatocyte function. Around 90% of the cells were competent for low-density lipoprotein (LDL) uptake (Fig. 3e, f) and accumulated glycogen (Fig. 3g). Abundant amounts of albumin were secreted into the medium (Fig. 3h), while hepatocyte cytochrome p450 function was induced (Fig. 3i). Yet, these *in vitro* functions remained less pronounced than those of freshly isolated hepatocytes.

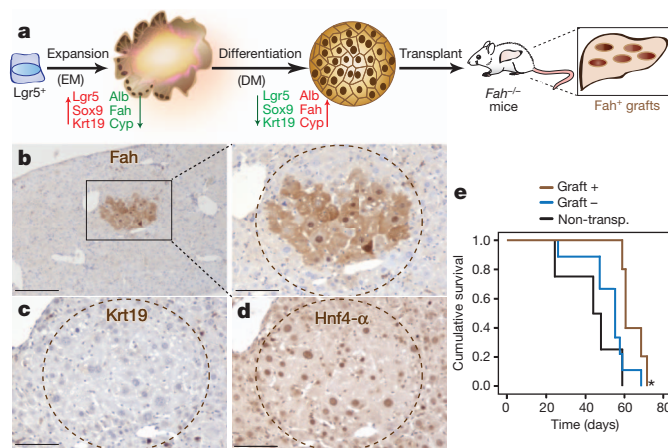
We then transplanted organoids from three independent clones into fumarylacetoacetate hydrolase (*Fah*<sup>-/-</sup>) mutant mice, a model for tyrosinemia type I liver disease. *Fah* deficiency results in liver failure unless the mutant mice are administered NTBC (2-(2-nitro-4-trifluoromethylbenzyl)-1,3-cyclohexanedione)<sup>30</sup>. Organoids derived from single *Lgr5*<sup>+</sup> cells expanded in expansion medium were cultured



**Figure 3 | Single-cell-derived hepatic organoids acquire hepatocyte fate and display hepatocyte functions *in vitro*.** **a–i**, Clonal *Lgr5*-derived cultures were grown in expansion medium and transferred to differentiation medium for 8–14 days before analysis. Two independent clonal cultures were analysed. **a–d**, Confocal images (z-stack projection) for hepatocyte-specific markers. **a**, Hnf4-α (red), albumin (Alb) (green) and EpCAM (grey). **b**, Mrp4 (green). **c**, ZO-1 (magenta). **d**, Alb (red), Krt19 (magenta) and hepatocyte surface marker OC2-2F8 (green, full description of OC2-2F8 marker in Supplementary Fig. 8). Nuclei were counterstained with Hoechst (blue). Scale bar, 50 μm (**a**, **d**) and 25 μm (**b**, **c**). **e**, **f**, Low-density lipoprotein (LDL) uptake was analysed using Dil-ac-LDL fluorescent substrate (red) in cultures maintained in expansion medium (EM) (**e**) or differentiation medium (DM) (**f**) for 14 days. Only cultures maintained in differentiation medium incorporated the substrate (red). Nuclei were counterstained with DRAQ5 (blue). Scale bar, 50 μm. **g**, Glycogen accumulation was determined by periodic acid-Schiff (PAS) staining in organoids grown in expansion medium or differentiation medium for 10 days. Graph shows the percentage of cells weakly or strongly positive for PAS. Results are shown as mean ± s.e.m. of 10 independent sections of 10 expansion medium or 10 differentiation medium independent organoids. Original magnification ×20. **h**, Albumin (Alb) secretion was measured in the supernatant collected for 24 h from clonal cultures maintained in expansion medium for 13 days or differentiation medium for 8 days (DM-d8) or 13 days (DM-d13). Results are expressed as picograms of albumin per cell. Mouse embryonic fibroblasts (MEF); hepatocytes (Hep.). **i**, Cyp3a activity was measured in cultures kept in differentiation medium for 13 days. Results are expressed as relative light units (RLU) per ml per million cells. **h**, **i**, MEFs and hepatocytes and HepG2 cells were used as negative and positive controls, respectively. Triplicates for each condition were analysed. Results are shown as mean ± s.e.m. of three independent experiments. \*\*\*, *P* < 0.0001.

in differentiation medium for 9 days and cell suspensions were injected intrasplenically in the mice (Fig. 4a). At 2 to 3 months post-transplantation, we analysed engraftment by *Fah* staining on serial sections of the entire liver of 15 recipient mice. We found *Fah*<sup>+</sup> nodules (Fig. 4b and Supplementary Fig. 9a–d), which occupied ~1% of the liver parenchyma, in two out of five mice transplanted with clone I and two out of five mice transplanted with clone III. For clone II, we only detect *Fah*<sup>+</sup> hepatocytes in one out of five mice analysed, in an area that occupied 0.1% of the total liver volume.

The histological results were confirmed by PCR analysis for the donor-cell gene *lacZ* (Supplementary Fig. 9e). Histological analysis revealed that the *Fah*<sup>+</sup> patches consisted of cells with hepatocyte morphology, Hnf4-α positive and Krt19 negative cells (Fig. 4c, d). This indicated that *in vivo* the cells had acquired a fully mature hepatocyte phenotype, while silencing any remaining ductal expression. No fusion of organoid cells with host hepatocytes was observed (Supplementary Fig. 9f). For comparison, transplantation of oval cells (MIC1-1C3<sup>+</sup>/133<sup>+</sup>/26<sup>+</sup>/CD45<sup>+</sup>/CD11b<sup>+</sup>/CD31<sup>+</sup>) has only resulted in trace engraftment (<0.1% of the total liver) in two out of twenty *Fah*<sup>-/-</sup> mice<sup>10</sup>. We compared the survival rate of the engrafted mice (graft<sup>+</sup>) to the non-engrafted mice (transplanted mice, negative for *Fah* staining, graft<sup>-</sup>) and to the non-transplanted controls (Fig. 4e). We observed a significant increase in survival of the graft<sup>+</sup> group compared to the graft<sup>-</sup> group (log rank = 0.02) and to the non-transplanted group (log rank = 0.007), indicating that the transplanted cells contributed to liver function *in vivo*. Unlike typical results obtained upon transplantation of freshly isolated hepatocytes, in which >30% of the liver repopulates and functional rescue is near 100%<sup>30</sup>, we did not observe a full rescue of the enzymatic defect, in concordance with the limited contribution of the transplanted areas to the overall liver volume. Such rescue will depend on further optimization of the differentiation and transplantation protocols. Competitive transplantation assays comparing normal hepatocytes to



**Figure 4 | Hepatocyte islands upon transplantation of clonal liver organoids into *Fah*<sup>-/-</sup> mutant mice.** *Lgr5*<sup>+</sup> single-cell derived liver organoids from 3 independent clones were expanded in the presence of expansion medium and differentiated for 9 days before transplantation into *Fah*<sup>-/-</sup> *Rag2*<sup>-/-</sup> *Il2rg*<sup>-/-</sup> (FRG) mice. Clone I (*n* = 8), clone II (*n* = 6), clone III (*n* = 5). **a**, Scheme showing the transplantation protocol. **b–d**, Representative positive graft within the liver parenchyma. **b**, *Fah*<sup>+</sup> area (clone III). The grafted cells were negative for Krt19 (**c**, ductal marker) and positive for Hnf4-α (**d**, hepatocyte marker). Scale bars, 400 μm (**b**, left panel) and 100 μm (**b**, right panel, **c**, **d**). **e**, Kaplan-Meier survival curve of transplanted mice with positive engraftment (brown curve, graft<sup>+</sup>, *n* = 5), transplanted mice without evidence of engraftment (blue curve, graft<sup>-</sup>, *n* = 9) and non-transplanted control mice (black curve, non-transp., *n* = 4). Plot displays the cumulative survival on a linear scale with 1.0 representing 100% survival. Kaplan-Meier survival analysis compares overall survival rates between two groups. Log-rank test is used to compare differences in survival. \*, log-rank = 0.02 (graft<sup>+</sup> versus graft<sup>-</sup>), log-rank = 0.007 (graft<sup>+</sup> versus non-transp.).



Lgr5-derived cells may reveal further phenotypic characteristics of the latter. Importantly, no dysplastic or anaplastic growth was detected in any of the recipient mice.

We report that damage of adult liver results in the expression of *Lgr5* in small cells near bile ducts. By lineage tracing, we demonstrated that these cells generate significant numbers of hepatocytes and biliary duct cells during the repair phase. The small *Lgr5*<sup>+</sup> cells express multiple Wnt target genes and other markers of intestinal *Lgr5*<sup>+</sup> stem cells. Yet, they carry the hallmarks of bipotent liver progenitors. Thus, *Lgr5* not only marks Wnt-driven stem cells that drive constitutive (intestine, stomach) or intermittent (hair follicle) physiological tissue self-renewal, but also defines a class of stem cells or progenitors that is called into action upon tissue damage. The Wnt-driven regenerative response can be exploited *in vitro* to expand freshly isolated duct fragments or even single *Lgr5*<sup>+</sup> cells into transplantable organoids. The *Rspo1*–*Lgr5* axis is crucial to the long-term growth and the observed genetic and phenotypic stability of the resulting organoids. Thus, the *Rspo1*–*Lgr5* axis allows adult stem cells to expand extensively in culture, like embryonic stem cells do. Our observations may serve as the basis for the development of regenerative strategies using adult stem or progenitor cells obtained from solid organs. Because these approaches can be based on the *in vitro* expansion of a single adult *Lgr5* progenitor cell, specific and safe genetic modifications may become feasible.

## METHODS SUMMARY

**Mouse experiments.** Animal experiments were performed in accordance with the institutional review committee at Hubrecht Institute and Oregon Health & Science University (IACUC protocol number IS00000119). Generation and genotyping of the *Lgr5-lacZ* and *Fah*<sup>−/−</sup> *Rag2*<sup>−/−</sup> *Il2rg*<sup>−/−</sup> (FRG) mice have been previously described<sup>1,30</sup>. *Axin2-LacZ* mice were obtained from European Mouse Mutant Archive. To induce liver injury, 3–5-month-old *Lgr5-lacZ*, *Axin2-lacZ*, *Lgr5-IRES-creERT2* × *Rosa26-lacZ* offspring or wild-type littermate BL6/BALB/c F<sub>1</sub> mice received an intraperitoneal (i.p.) injection of CCl<sub>4</sub> (1 ml kg<sup>−1</sup>, Sigma-Aldrich) dissolved in corn oil, or corn oil alone. *Lgr5-IRES-creERT2* × *Rosa26-lacZ* mice were fed with a diet supplemented with 0.1% (w/w) DDC or supplemented with MCD and 0.1% ethionine in the drinking water (MCDE) or with regular diet (not supplemented). Four or five days after liver injury, lineage tracing was induced by one or two i.p. injections of tamoxifen (3 mg per mouse). Detailed lineage tracing protocols are provided in the Methods.

**Liver organoid culture.** Biliary ducts or sorted *Lgr5-lacZ*<sup>+</sup> cells were isolated, mixed with Matrigel (BD Bioscience) and cultured as described in ref. 2. Medium composition was AddMEM/F12 (Invitrogen) supplemented with B27 and N2 (Invitrogen), *N*-acetylcysteine (1.25 μM, Sigma-Aldrich), gastrin (10 nM, Sigma), EGF (50 ng ml<sup>−1</sup>, Peprotech), 10% RSP01 conditioned medium (provided by C. Kuo), Fgf10 (100 ng ml<sup>−1</sup>, Peprotech), nicotinamide (10 mM, Sigma-Aldrich) and HGF (50 ng ml<sup>−1</sup>, Peprotech). Detailed protocols are provided in the Methods.

**Transplantation assay.** Three *Lgr5*-derived single cell clones (I–III) were grown for at least 3 months and differentiated for 9 days before transplantation to *Fah*<sup>−/−</sup> *Rag2*<sup>−/−</sup> *Il2rg*<sup>−/−</sup> mice (intrasplenic injection). Detailed protocols are provided in the Methods.

**Full Methods** and any associated references are available in the online version of the paper.

Received 10 September 2011; accepted 7 December 2012.

Published online 27 January 2013.

1. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
2. Barker, N. *et al.* *Lgr5*<sup>+</sup> stem cells drive self-renewal in the stomach and build long-lived gastric units *in vitro*. *Cell Stem Cell* **6**, 25–36 (2010).
3. Jaks, V. *et al.* *Lgr5* marks cycling, yet long-lived, hair follicle stem cells. *Nature Genet.* **40**, 1291–1299 (2008).
4. Sato, T. *et al.* Single *Lgr5* stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
5. Yui, S. *et al.* Functional engraftment of colon epithelium expanded *in vitro* from a single adult *Lgr5*<sup>+</sup> stem cell. *Nature Med.* **18**, 618–623 (2012).

6. Kim, K. A. *et al.* Mitogenic influence of human R-spondin1 on the intestinal epithelium. *Science* **309**, 1256–1259 (2005).
7. Carmon, K. S., Gong, X., Lin, Q., Thomas, A. & Liu, Q. R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/β-catenin signaling. *Proc. Natl Acad. Sci. USA* **108**, 11452–11457 (2011).
8. de Lau, W. *et al.* *Lgr5* homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* **476**, 293–297 (2011).
9. Duncan, A. W., Dorrell, C. & Grompe, M. Stem cells and liver regeneration. *Gastroenterology* **137**, 466–481 (2009).
10. Dorrell, C. *et al.* Prospective isolation of a bipotential clonogenic liver progenitor cell in adult mice. *Genes Dev.* **25**, 1193–1203 (2011).
11. Sackett, S. D. *et al.* Foxl1 is a marker of bipotential hepatic progenitor cells in mice. *Hepatology* **49**, 920–929 (2009).
12. Furuyama, K. *et al.* Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine. *Nature Genet.* **43**, 34–41 (2011).
13. Shin, S. *et al.* Foxl1-Cre-marked adult hepatic progenitors have clonogenic and bilineage differentiation potential. *Genes Dev.* **25**, 1185–1192 (2011).
14. Benhamouche, S. *et al.* Apc tumor suppressor gene is the 'zonation-keeper' of mouse liver. *Dev. Cell* **10**, 759–770 (2006).
15. Hu, M. *et al.* Wnt/β-catenin signaling in murine hepatic transit amplifying progenitor cells. *Gastroenterology* **133**, 1579–1591 (2007).
16. Lustig, B. *et al.* Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. *Mol. Cell. Biol.* **22**, 1184–1193 (2002).
17. Stowell, R. E. & Lee, C. S. Histochemical studies of mouse liver after single feeding of carbon tetrachloride. *AMA Arch. Pathol.* **50**, 519–537 (1950).
18. Muñoz, J. *et al.* The *Lgr5* intestinal stem cell signature: robust expression of proposed quiescent '4' cell markers. *EMBO J.* **31**, 3079–3091 (2012).
19. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).
20. Akhurst, B. *et al.* A modified choline-deficient, ethionine-supplemented diet protocol effectively induces oval cells in mouse liver. *Hepatology* **34**, 519–522 (2001).
21. Preisegger, K. H. *et al.* Atypical ductular proliferation and its inhibition by transforming growth factor beta1 in the 3,5-diethoxycarbonyl-1,4-dihydrocollidine mouse model for chronic alcoholic liver disease. *Lab. Invest.* **79**, 103–109 (1999).
22. Michalopoulos, G. K., Bowen, W. C., Mulé, K. & Stolz, D. B. Histological organization in hepatocyte organoid cultures. *Am. J. Pathol.* **159**, 1877–1887 (2001).
23. Schmelzer, E. *et al.* Human hepatic stem cells from fetal and postnatal donors. *J. Exp. Med.* **204**, 1973–1987 (2007).
24. Kamiya, A., Kakinuma, S., Yamazaki, Y. & Nakauchi, H. Enrichment and clonal culture of progenitor cells during mouse postnatal liver development in mice. *Gastroenterology* **137**, 1114–1126 (2009).
25. Duncan, A. W. *et al.* The ploidy conveyor of mature hepatocytes as a source of genetic variation. *Nature* **467**, 707–710 (2010).
26. Tanimizu, N. & Miyajima, A. Notch signaling controls hepatoblast differentiation by altering the expression of liver-enriched transcription factors. *J. Cell Sci.* **117**, 3165–3174 (2004).
27. Lemaigre, F. P. Mechanisms of liver development: concepts for understanding liver disorders and design of novel therapies. *Gastroenterology* **137**, 62–79 (2009).
28. Zaret, K. S. Genetic programming of liver and pancreas progenitors: lessons for stem-cell differentiation. *Nature Rev. Genet.* **9**, 329–340 (2008).
29. Si-Tayeb, K., Lemaigre, F. P. & Duncan, S. A. Organogenesis and development of the liver. *Dev. Cell* **18**, 175–189 (2010).
30. Azuma, H. *et al.* Robust expansion of human hepatocytes in *Fah*<sup>−/−</sup> *Rag2*<sup>−/−</sup> *Il2rg*<sup>−/−</sup> mice. *Nature Biotechnol.* **25**, 903–910 (2007).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank H. Begthel, A. Bujs, W. Karthaus, C. Kroon-Veenboer, M. van den Born, S. van der Brink, and L. Zeinstra for technical assistance. This work was supported by grants to M.H. (EU/236954), S.F.B. (EU/232814), and V.S.W.L. and J.H.v.E. (Ti Pharma/T3-106).

**Author Contributions** Experiments were conceived and designed by M.H. and H.C. Transplantation experiments were conceived and designed by M.H., C.D., M.G. and H.C. Experiments were performed by M.H., C.D., S.F.B. and V.S.W.L. A.H. helped with the transplantation experiments. M.v.d.W. and N.S. performed the sortings and K.H. the karyotyping experiments. J.H.v.E. designed and generated the *Lgr5-IRES-creERT2* allele. V.S.W.L. performed the bioinformatic analysis of the microarrays and M.J.F. the Y-chromosome staining. M.H. and T.S. developed the isolation protocol. T.S. helped in the beginning phase of the project. R.G.V. helped with sorting experiments. M.H. analysed the data. M.H. and H.C. wrote the manuscript. The other authors commented on the manuscript.

**Author Information** The data for the microarray analysis have been deposited to the Gene Expression Omnibus under the accession number GSE32210. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.C. ([h.clevers@hubrecht.eu](mailto:h.clevers@hubrecht.eu)).



## METHODS

**Tracing experiment.** *Lgr5-IRES-creERT2* mice were crossed with the *Rosa26-lacZ* Cre reporter mice<sup>19</sup>. Offspring that were 8–12-weeks-old received tamoxifen (3 mg), and the mice were analysed at day 9 and day 45 post-induction. Following validation of the new allele, *Lgr5-IRES-creERT2* × *Rosa26-lacZ* offspring were given an intraperitoneal injection of 1 ml kg<sup>-1</sup> CCl<sub>4</sub> diluted in corn oil or corn oil alone and 5 days later received a single dose of tamoxifen (3 mg per mouse). Mice were killed 2, 4, 7 and 21 days later and tissue was analysed for the presence of tracing events. To evaluate the oval cell damage response, *Lgr5-IRES-creERT2* × *Rosa26-lacZ* offspring were fed with a DDC or MCDE diet and 4 and 6 days later given a 3 mg i.p. injection of tamoxifen. For ethical reasons, 2 days after the last dose of tamoxifen (day 8 after diet initiation), the mice were returned to normal diet until death. Mice were killed 9 and 21 days after the first tamoxifen injection.

**Liver organoid cell culture and single cell culture.** Biliary ducts were isolated from the bulk of the tissue by collagenase-aided dissociation (Collagenase type XI 0.012%, dispase 0.012%, FBS 1% in DMEM medium). Isolated ducts were mixed with Matrigel (BD Bioscience) and seeded and cultured as described previously<sup>2,4</sup>. After Matrigel formed a gel, culture medium was added. Culture medium was based on AddMEM/F12 (Invitrogen) supplemented with B27 and N2 (both Invitrogen), 1.25 μM *N*-acetylcysteine (Sigma-Aldrich), 10 nM gastrin (Sigma-Aldrich) and the following growth factors: 50 ng ml<sup>-1</sup> EGF (Peprotech), 1 μg ml<sup>-1</sup> Rspo1 (Nuvelo) or 10% RSP01 conditioned medium (provided by C. Kuo and prepared as in ref. 31), 100 ng ml<sup>-1</sup> Fgf10 (Peprotech), 10 mM nicotinamide (Sigma-Aldrich) and 50 ng ml<sup>-1</sup> HGF (Peprotech). For the first 4 days after seeding, the cells were also supplemented with Noggin (100 ng ml<sup>-1</sup>) and Wnt3a-conditioned medium (prepared as in ref. 2). One week after seeding, organoids were removed from the Matrigel, mechanically dissociated into small fragments, and transferred to fresh Matrigel. Passage was performed in 1:4–1:8 split ratio once per week for more than 12 months.

For clonogenic assays, single-cell suspensions from either established cultures or from liver tissue from *Lgr5-lacZ* mice treated with corn oil or *Lgr5-lacZ* mice and wild-type littermates injected with CCl<sub>4</sub> were stained with DetectaGene green CMFDG (5-chloromethylfluorescein di-β-D-galactopyranoside) LacZ gene expression kit (Molecular Probes) according to the manufacturer's instructions. Wild-type littermates and *Lgr5-lacZ* littermates treated with corn oil only were used to facilitate cell subfractioning and exclusion of nonspecific staining. Propidium iodide staining was used to label dead cells and FSC pulse-width gating to exclude cell doublets (MoFlow, Dako). Sorted *Lgr5-lacZ* cells were embedded in Matrigel and seeded in 96-well plates at a ratio of one cell per well. Cells were cultured as described above with medium supplemented with Y-27632 (10 μM, Sigma-Aldrich) for the first 4 days. Passage was performed in split ratios of 1:4–1:8 once per week for at least 8 months. All phase-contrast images were acquired using a Leica DMIL microscope and a Leica DFC420C camera.

**Hepatocyte differentiation.** To enhance hepatocyte cell fate, *Lgr5-lacZ* single-cell-derived liver organoids were seeded and kept 2–4 days under the liver expansion conditions described previously. Then the medium was changed to no longer contain Rspo1, HGF and nicotinamide and instead to contain EGF (50 ng ml<sup>-1</sup>), Fgf10 (100 ng ml<sup>-1</sup>), A8301 (50 nM, Tocris Bioscience) and DAPT (10 nM, Sigma-Aldrich). For transplantation and *in vitro* functional studies, cultures were also supplemented with dexamethasone (30 μM) for the last 3 days of the differentiation. Medium was changed every other day for a period of 9 to 14 days.

**β-galactosidase (LacZ) staining, immunohistochemistry and immunofluorescence.** Tissues were fixed for 2 h in ice-cold fixative (1% formaldehyde; 0.2% glutaraldehyde; 0.02% NP40 in phosphate buffered saline (PBS) and incubated overnight at room temperature with 1–2 mg ml<sup>-1</sup> of X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) solution as described in ref. 2. The stained tissues were transferred to tissue cassettes and paraffin blocks were prepared using standard methods. Tissue sections (4 μM) were prepared and counterstained with neutral red. For immunohistochemistry, tissues and organoids were fixed using 4% formaldehyde, and stained using standard histology techniques as described in ref. 2. The antibodies and dilutions used are listed in Supplementary Table 4. Stained tissues were counterstained with Mayer's hematoxylin. Pictures were taken with a Nikon E600 camera and a Leica DFDC500 microscope. For whole-mount immunofluorescent staining, organoids were processed as described in ref. 2. Nuclei were stained with Hoechst 33342 (Molecular Probes). Images were acquired using a confocal microscope (Leica, SP5).

**Karyotyping of liver cultures.** Organoid cultures in exponential growing phase were incubated for 16 h with 0.05 μg ml<sup>-1</sup> colcemid (Gibco). Then cultures were washed and dissociated into single cells using TrypLE express (Gibco) and processed as described in ref. 25. Chromosomes from 100 metaphase-arrested cells were counted.

**Hepatocyte functional studies.** To assess glycogen storage and LDL uptake, liver organoids grown in expansion medium or differentiation medium for 14 days

were stained by periodic acid-Schiff (PAS, Sigma) and DiI-Ac-LDL (Biomedical Technologies), respectively, following the manufacturer's instructions. To determine albumin secretion, liver organoids were differentiated as described. Culture medium was changed every other day and culture supernatant was collected at days 8 or 13 after differentiation started. In all cases culture medium was collected 24 h after the last medium change. Isolated hepatocytes (by classical collagenase perfusion) and MEFs were cultured for 24 h in the same medium without growth factors and were used as positive and negative controls, respectively. The amount of albumin in culture supernatant was determined using a mouse-specific albumin ELISA kit (Assaypro). To measure Cyp3a activity the cultures were differentiated for 12–14 days as described. On the day of the experiment, cells were removed from the Matrigel and transferred to non-attaching plates and cultured with the Luciferin-PFBE substrate (50 μM) in hepatocyte medium supplemented with 10% FBS (Gibco). As controls, hepatocytes (isolated from BL6 mice by classical collagenase perfusion) HepG2 cells (ATCC) and MEFs were cultured for 24 h in DMEM 10% FBS and on the day of the experiment transferred to hepatocyte medium supplemented with 10% FBS (Gibco) and luciferin-PFBE substrate (50 μM). Cytochrome P450 activity was measured 8 h later using the P450-Glo assay kit (Promega) according to the manufacturer's instructions.

**Flow cytometry analysis.** Exponentially growing organoids derived from a single *Lgr5-lacZ* cell were cultured for 9 days in differentiation medium. Then organoids were dissociated into single cells, resuspended in DMEM plus 2% fetal bovine serum (FBS) and incubated with OC2-2F8 hybridoma supernatant (described in Supplementary Fig. 8) for 30 min. APC-conjugated goat anti-rat immunoglobulin adsorbed against mouse serum proteins was used as a secondary antibody (Jackson ImmunoResearch). Cells were analysed with a Cytopeia inFluxV-GS (Becton Dickinson); for FSC, propidium iodide was used to label dead cells for exclusion and pulse-width gating used to exclude cell doublets.

**Microarray analysis.** For the expression analysis, total RNA was isolated from cultured organoids in expansion medium or differentiation medium and from mouse adult liver, pancreas, newborn white adipose tissue, brown adipose tissue, muscle, FACS-sorted *Lgr5*<sup>+</sup> cells, ductal cells and hepatocytes using RNAase kit (Qiagen) following the manufacturer's instructions. From 50 ng to 1,000 ng of total RNA were labelled with low RNA Input Linear Amp kit (Agilent). Universal mouse reference RNA (Agilent) was differentially labelled and hybridized to the tissue or cultured samples. A 4 × 44 K whole mouse genome dual colour microarray (G4122F, Agilent) was used. Labelling, hybridization and washing were performed according to Agilent guidelines. Microarray signal and background information were retrieved using the Feature Extraction program (V.9.5.3, Agilent). Normalized data were analysed by hierarchical clustering using Cluster 3 and visualized with TreeView. A filter of >threefold difference in at least two arrays was chosen for CCl<sub>4</sub> liver treated microarray (Fig. 1c) and *Lgr5*<sup>+</sup> cell sorted microarray (Fig. 1g). A filter of greater than fourfold difference in at least two arrays was chosen for all organoid (EM and DM) microarrays analysed. The data for the microarray analysis have been deposited to the Gene Expression Omnibus under the accession number GSE32210.

**qRT-PCR.** RNA was extracted from cell cultures or freshly isolated tissue using the RNeasy Mini RNA Extraction Kit (Qiagen) and reverse-transcribed using Moloney murine leukaemia virus reverse transcriptase (Promega). cDNA was amplified in a thermal cycler (GeneAmp PCR System 9700, Applied Biosystems) as described in ref. 32. Primers used are listed in Supplementary Table 5.

**Transplantation assay.** *FRG* mice were maintained with 16 mg l<sup>-1</sup> of 2-(2-nitro-4-trifluoro-methyl-benzoyl)-1,3 cyclohexanedione (NTBC) in drinking water as described in ref. 30. For transplantation, three clones (clone I, II and III) derived from two different *Lgr5-lacZ* mice were grown for at least 3 months. Cultures were kept in liver expansion medium and transferred to differentiation medium 9 days before transplantation. Suspensions of 500, 000 to 800,000 organoid-derived cells were injected intrasplenically to *Fah*<sup>-/-</sup> *Rag2*<sup>-/-</sup> *IL2rg*<sup>-/-</sup> mice. Kineret/Anakinra (Biovitrum) was added to the cellular suspension at the time of transplant to prevent immunorejection. Mice were given the NTBC drug in drinking water for 4 days following transplantation. Then NTBC was removed and mouse health status and body weight were monitored every other day. Two weeks later, one mouse (clone III) was killed for engraftment analysis purposes. The remaining mice underwent a second round of selection (NTBC was re-administered) after which NTBC was permanently discontinued. For each mouse the end point was set to 20% loss in body weight or when indicated for ethical reasons. *FRG* littermates injected with freshly isolated hepatocytes or non-transplanted littermates were used as positive and negative controls, respectively.

**Y-chromosome fluorescent *in situ* hybridization staining.** Y-chromosome staining was performed to exclude fusion of the clones derived from female donor cells into male recipients by using standard FISH techniques as described elsewhere<sup>33</sup>.

**Statistics.** All data are presented as mean ± s.e.m. Mann-Whitney non-parametric test was used. *P* < 0.05 was considered statistically significant. In all cases data

from at least three independent experiments was used. All calculations were performed using SPSS software package.

Survival analysis was performed to analyse time-to-event probability. The survival curves of the three groups (1) engrafted, transplanted mice, (2) non-engrafted, transplanted mice and (3) non-transplanted controls were compared. Log-rank test was used to determine the statistical significance of the differences in time-to-event. Log rank  $<0.05$  was considered statistically significant. All calculations were performed using SPSS software package.

31. Farin, H. F., Van Es, J. H. & Clevers, H. Redundant sources of Wnt regulate intestinal stem cells and promote formation of paneth cells. *Gastroenterology* **143**, 1518–1529 (2012).
32. Huch, M., Gros A., José A., González J. R., Alemnay R. & Fillat C.. Urokinase-type plasminogen activator receptor transcriptionally controlled adenoviruses eradicate pancreatic tumors and liver metastasis in mouse models. *Neoplasia* **11**, 518–528 (2009).
33. Wang, X. *et al.* Cell fusion is the principal source of bone-marrow-derived hepatocytes. *Nature* **422**, 897–901 (2003).

# Modelling vemurafenib resistance in melanoma reveals a strategy to forestall drug resistance

Meghna Das Thakur<sup>1</sup>, Fernando Salangsang<sup>1</sup>, Allison S. Landman<sup>2</sup>, William R. Sellers<sup>3</sup>, Nancy K. Pryer<sup>1</sup>, Mitchell P. Levesque<sup>4</sup>, Reinhard Dummer<sup>4</sup>, Martin McMahon<sup>2</sup> & Darrin D. Stuart<sup>1</sup>

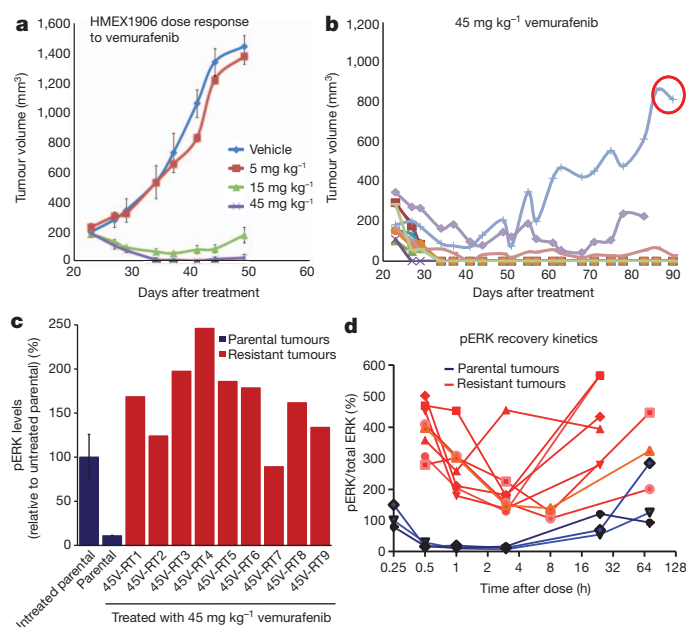
Mutational activation of *BRAF* is the most prevalent genetic alteration in human melanoma, with  $\geq 50\%$  of tumours expressing the *BRAF*(V600E) oncoprotein<sup>1,2</sup>. Moreover, the marked tumour regression and improved survival of late-stage *BRAF*-mutated melanoma patients in response to treatment with vemurafenib demonstrates the essential role of oncogenic *BRAF* in melanoma maintenance<sup>3,4</sup>. However, as most patients relapse with lethal drug-resistant disease, understanding and preventing mechanism(s) of resistance is critical to providing improved therapy<sup>5</sup>. Here we investigate the cause and consequences of vemurafenib resistance using two independently derived primary human melanoma xenograft models in which drug resistance is selected by continuous vemurafenib administration. In one of these models, resistant tumours show continued dependency on *BRAF*(V600E)→MEK→ERK signalling owing to elevated *BRAF*(V600E) expression. Most importantly, we demonstrate that vemurafenib-resistant melanomas become drug dependent for their continued proliferation, such that cessation of drug administration leads to regression of established drug-resistant tumours. We further demonstrate that a discontinuous dosing strategy, which exploits the fitness disadvantage displayed by drug-resistant cells in the absence of the drug, forestalls the onset of lethal drug-resistant disease. These data highlight the concept that drug-resistant cells may also display drug dependency, such that altered dosing may prevent the emergence of lethal drug resistance. Such observations may contribute to sustaining the durability of the vemurafenib response with the ultimate goal of curative therapy for the subset of melanoma patients with *BRAF* mutations.

To model the emergence of drug resistance, we developed an early passage, vemurafenib-naïve, primary human-patient-derived xenograft (PDX) *BRAF*<sup>T1799A</sup>-mutated melanoma model, HMEX1906 (Supplementary Table1), which was continuously treated with vemurafenib in immunocompromised mice. This system models the emergence of drug-resistant melanoma in response to drug exposures similar to those in patients. Furthermore, this model permits the sampling of serial biopsies from a single tumour, allowing us to investigate the presence of more than one clonally derived mechanism of resistance within the original tumour.

HMEX1906 melanomas are highly sensitive to vemurafenib, with tumour regression observed at clinically relevant drug exposures (Fig. 1a and Supplementary Fig. 1a–c). To generate drug-resistant melanomas, tumour-bearing mice were dosed for 8 weeks with 45 mg kg<sup>-1</sup> vemurafenib. This dose resulted in over 80% inhibition of phosphorylated (p) ERK1 and ERK2 (also known as MAPK3 and MAPK1, respectively) (Supplementary Fig. 1d) for up to 24 h, a degree of inhibition previously associated with tumour regression in clinical trials<sup>4,5</sup>. Approximately 56 days after dosing was initiated, drug-resistant tumours emerged in 2 out of 10 mice (Fig. 1b). One such tumour (45V-RT) was harvested, fragmented and re-implanted into a new cohort of mice, which were then treated with 45 mg kg<sup>-1</sup> vemurafenib

to generate drug-resistant tumours for exploration of mechanisms of resistance (Supplementary Fig. 1e, f).

Next, we assessed differences in the response to vemurafenib between sensitive parental HMEX1906 and resistant 45V-RT tumours by measuring pERK1 and pERK2 levels 3 h after drug dosing (Fig. 1c). Whereas pERK1/2 and the expression of ERK1/2 target genes such as *DUSP6* and *SPRY4* were strongly suppressed in sensitive parental HMEX1906 tumours, they were largely unaffected in drug-resistant 45V-RT tumours (Fig. 1c and Supplementary Fig. 2a, b). Analysis of fine needle aspirates (FNAs) of eight resistant tumours over a 72-h time course revealed higher pERK1/2 levels compared to parental tumours 30 min after drug administration, with the nadir of pERK1/2 consistently higher than that observed in parental drug-sensitive tumours



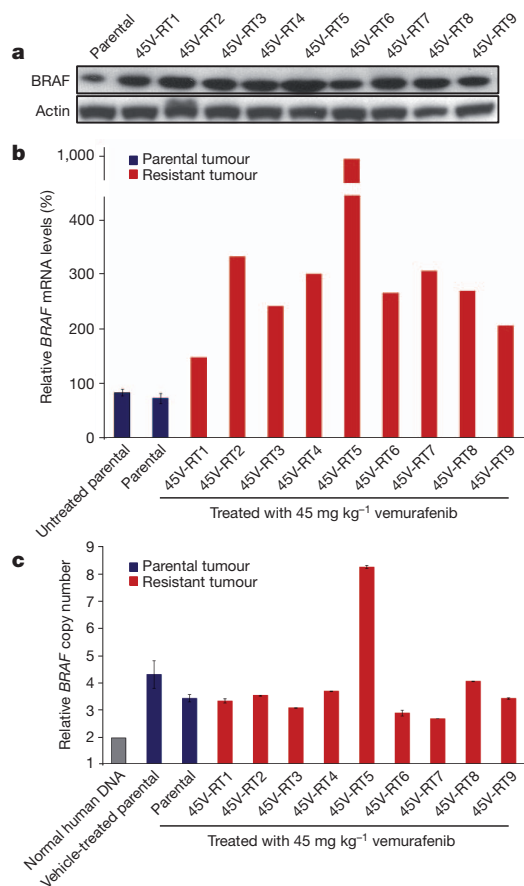
**Figure 1 | Resistance to vemurafenib in a primary human melanoma xenograft model.** **a**, Mice bearing subcutaneous HMEX1906 tumours were dosed with vehicle ( $n = 10$ ), 5 mg kg<sup>-1</sup> ( $n = 8$ ), 15 mg kg<sup>-1</sup> ( $n = 8$ ) or 45 mg kg<sup>-1</sup> ( $n = 10$ ) vemurafenib twice daily (mean tumour volume  $\pm$  s.e.m.). **b**, Continuous dosing of tumour-bearing mice over an extended time leads to the emergence of resistant tumours. The tumour circled in red was excised, subdivided and re-implanted to be used for further analysis. **c**, Parental tumours ( $n = 3$  untreated and treated, mean pERK levels  $\pm$  s.e.m. for the three different tumours) and resistant tumours were treated with 45 mg kg<sup>-1</sup> vemurafenib, and lysates were collected 3 h after the drug dose to measure pathway inhibition using pERK levels. **d**, The pharmacodynamics of pERK1/2 were evaluated over multiple time points for eight resistant tumours (red) and three parental tumours (blue).

<sup>1</sup>Novartis Institutes for Biomedical Research, Emeryville, California 94608, USA. <sup>2</sup>Helen Diller Family Comprehensive Cancer Center & Department of Cellular & Molecular Pharmacology, University of California San Francisco, California 94143-0128, USA. <sup>3</sup>Novartis Institutes for Biomedical Research, Cambridge, Massachusetts 02139, USA. <sup>4</sup>Department of Dermatology, University Hospital Zurich, Gloriastrasse 31, 8091 Zurich, Switzerland.



(Fig. 1d). Hence, resistant tumours do respond to drug treatment, but the degree of pERK1/2 inhibition was less profound compared to sensitive melanomas. These data suggest that BRAF(V600E) remains essential for sustaining MEK→ERK pathway activation. One explanation for such observations is that BRAF was mutated to a vemurafenib-resistant state. Alternatively, upstream (for example, *NRAS*) or downstream (for example, *MEK1*; also known as *MAP2K1*) nodes in the RAS→RAF→MEK→ERK pathway may be mutationally activated, as described recently<sup>6,7</sup>. However, exome sequence analysis failed to reveal secondary mutations in the coding sequences of *BRAF*, *NRAS*, *KRAS*, *HRAS* or *MEK1* in resistant tumours (data not shown).

To determine whether BRAF overexpression or alternative splicing might account for vemurafenib resistance<sup>8,9</sup>, BRAF(V600E) expression was measured in sensitive and resistant tumours. Immunoblot analysis indicated that both sensitive and resistant tumours expressed an 85 kilodalton (kDa) isoform of BRAF(V600E) (Fig. 2a). However, compared to sensitive tumours, all nine resistant tumours expressed elevated levels of BRAF messenger RNA and protein, with the 45V-RT5 tumour showing the highest levels (Fig. 2a, b). Taqman analysis of *BRAF* copy number indicated that the parental HMEX1906 tumour contained approximately six copies of *BRAF*<sup>T1799A</sup>. Although eight out of nine of the resistant tumours showed no additional *BRAF* copy number gain, the 45V-RT5 tumour was found to have ~14 copies of *BRAF*<sup>T1799A</sup> (Fig. 2c), consistent with *BRAF* amplification as a mechanism of vemurafenib resistance<sup>9</sup>. These data suggest that the

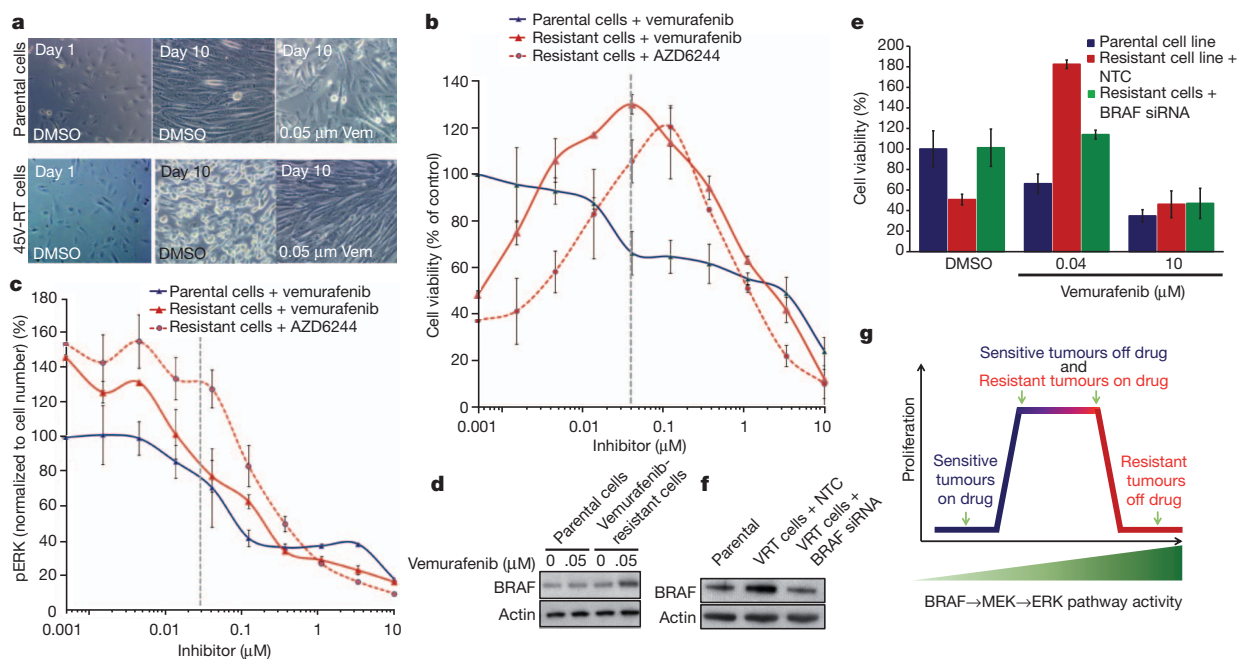


**Figure 2 | Resistant tumours show increased BRAF(V600E) expression.** **a**, BRAF protein level was determined by western blot (with actin as a loading control) in parental and resistant tumours (all lysates were collected 3 h after the drug dose). **b**, BRAF mRNA was measured by quantitative polymerase chain reaction with reverse transcription (RT-qPCR), ( $n = 3$  untreated and treated independent parental tumours, BRAF mRNA levels  $\pm$  s.e.m.). **c**, BRAF copy number was determined by qPCR of genomic DNA ( $n = 3$  untreated and treated independent parental tumours, BRAF copy number levels  $\pm$  s.e.m.).

parental tumour contains heterogeneous vemurafenib-resistant cells, all of which show elevated BRAF mRNA/protein expression but only a subpopulation further amplify *BRAF*<sup>10</sup>. Finally, we did not detect evidence of alternatively spliced isoforms of BRAF(T1799A) or BRAF(V600E) by mRNA or protein analysis<sup>8</sup>.

To confirm elevated levels of BRAF protein as a resistance mechanism, melanoma cell lines were derived from the parental HMEX1906 and the vemurafenib-resistant 45V-RT tumour (Fig. 1b). We noted difficulty in establishing cultures of drug-resistant cells unless the media contained ~50 nM vemurafenib. This observation is consistent with reports that vemurafenib-resistant variants of BRAF(V600E)-expressing M288, SK-MEL28 or M14 melanoma cell lines require vemurafenib for continuous proliferation<sup>11</sup>. In addition, HMEX1906 melanoma cells grown in the absence of drug and 45V-RT melanoma cells grown in the presence of drug showed similar morphology (Fig. 3a, top middle and bottom right). However, culturing 45V-RT melanoma cells in the absence of vemurafenib for 10 days resulted in marked alterations in cell morphology. Cells appeared rounded, refractile and spindle shaped, features characteristic of cells with elevated RAF→MEK→ERK signalling<sup>12–14</sup> (Fig. 3a, middle). Furthermore, a cell-proliferation assay conducted with a range of drug concentrations indicated a bell-shaped response to vemurafenib, with peak proliferation in the resistant cells occurring at 50 nM vemurafenib, and with diminished cell proliferation noted at lower and higher drug concentrations (Fig. 3b). A similar curve was observed with the MEK inhibitor AZD6244, with a shift in peak proliferation consistent with the compound's decreased potency. Analysis of BRAF(V600E)→MEK→ERK signalling indicated that reducing the concentration of vemurafenib led to elevated pERK1/2 levels in the resistant 45V-RT cells. Moreover, the level of pERK1/2 in resistant 45V-RT cells cultured in 50 nM vemurafenib was similar to that detected in parental HMEX1906 cells cultured in the absence of vemurafenib (Fig. 3c, dotted line). After 10 days of culture in the presence of 50 nM vemurafenib, 45V-RT cells showed elevated BRAF(V600E) protein expression similar to resistant tumours in mice (Figs 2a and 3d). To confirm that resistant cells remained oncogene dependent, we inhibited BRAF(V600E) expression by RNA interference (using siBRAF short interfering RNA). Complete knockdown of BRAF(V600E) expression in resistant 45V-RT cells resulted in suppression of proliferation; hence, resistant cells remain dependent on oncogenic BRAF(V600E) signalling for proliferation (Supplementary Fig. 3a, b). However, partial suppression of BRAF(V600E) to levels detected in parental cells (Fig. 3f) re-sensitized resistant cells to both vemurafenib and AZD6244 (Fig. 3e and Supplementary Fig. 3c). These data confirm that resistant tumour cells remain oncogene dependent and that drug resistance is due to elevated expression of BRAF(V600E). Moreover, the fitness benefit given to resistant cells by elevated BRAF(V600E) in the presence of vemurafenib becomes a fitness deficit when the drug is removed. To test this hypothesis, we expressed a conditional BRAF(V600E)-oestrogen receptor (ER; also known as ESR1) fusion protein in parental HMEX1906 cells, such that addition of 4-hydroxytamoxifen (4-HT) leads to increased BRAF(V600E) signalling<sup>12–14</sup> (Supplementary Fig. 4a). As predicted, elevated BRAF(V600E) activity in the parental cells led to increased pERK levels but decreased proliferation (Supplementary Fig. 4a, b). These data indicate that HMEX1906 cells are responding to both the quality and quantity of BRAF(V600E)→MEK→ERK signalling such that either reduced (in response to vemurafenib) or enhanced (in response to BRAF(V600E)-ER activation) pathway activation has a deleterious effect on their proliferation<sup>12–14</sup>.

To test whether observations made with cultured melanoma cells are relevant to tumorigenesis *in vivo*, we evaluated the effects of cessation of drug administration on vemurafenib-resistant tumours in mice. Initially we noted that significantly fewer drug-resistant tumours grew in vehicle-treated mice as compared to vemurafenib-treated mice (Fig. 4a). Furthermore, cessation of drug treatment of mice carrying vemurafenib-resistant melanomas led to clear signs of regression



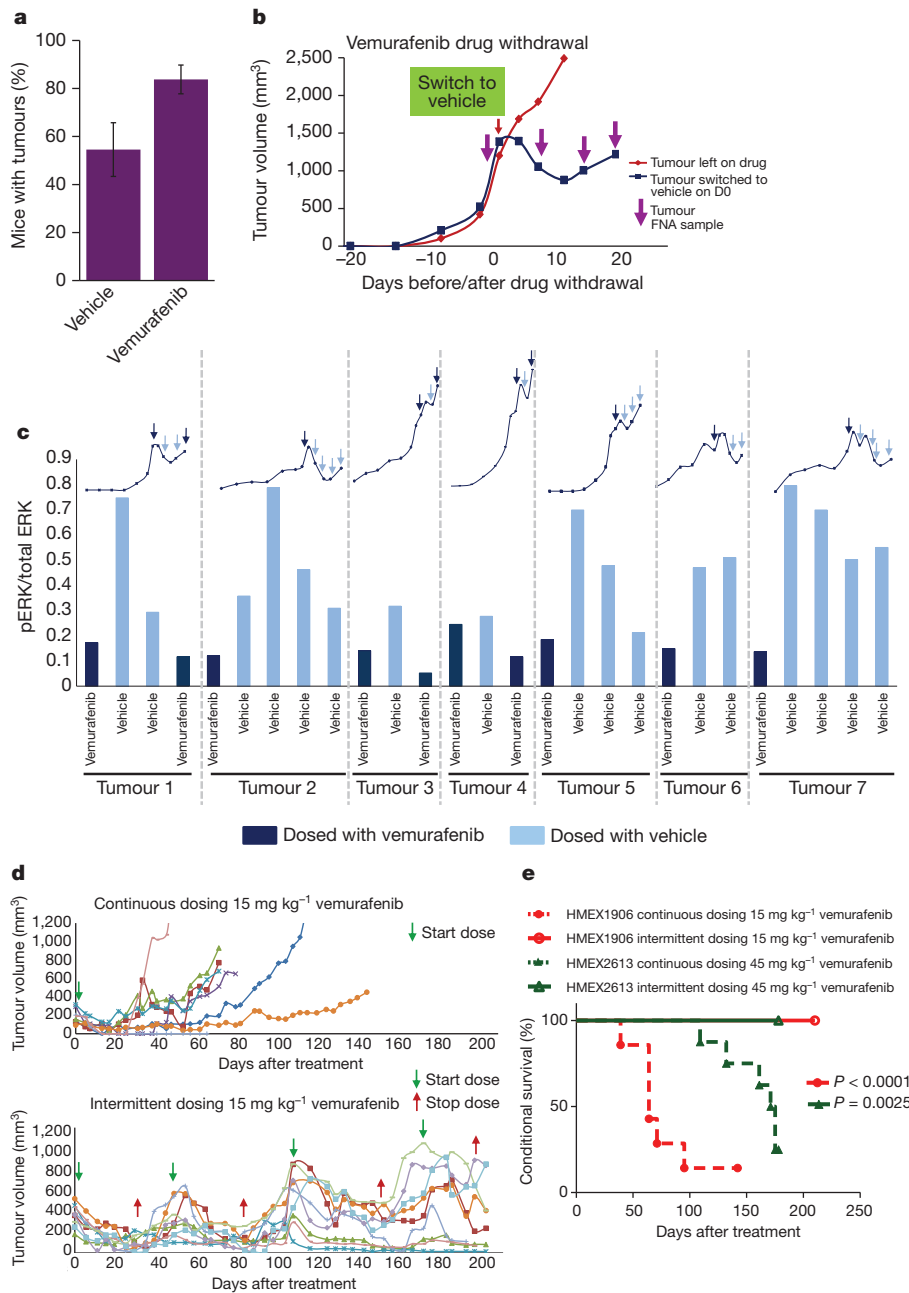
**Figure 3 | Vemurafenib-resistant tumour cells require continuous exposure to vemurafenib.** **a**, Parental (top) and vemurafenib-resistant (bottom) tumour-derived cells were imaged after 1 day (left), after 10 days (middle) in culture (0.05% dimethylsulphoxide (DMSO)), and after 10 days of culture in 0.05  $\mu$ M vemurafenib (Vem; right). Original magnification,  $\times 40$ . **b**, Parental and vemurafenib-resistant cells were treated with the indicated concentrations of vemurafenib and AZD6244 for 72 h, and viability was determined using the Cell Titer-Glo ATP-based luminescence assay, with DMSO-treated parental cells set as the control. **c**, A parallel plate similar to

**b** was set up and corresponding pERK1/2 levels were measured from samples. **d**, BRAF protein level was determined in parental and resistant tumour cells by western blot. **e**, Resistant and parental tumour cells were subjected to BRAF siRNA, treated with vemurafenib or control (DMSO), and cell viability was determined by Cell Titer-Glo assay after 3 days of culture. NTC, non-targeting control. **f**, BRAF knockdown efficiency was determined by western blot with actin as loading control. **g**, Model correlating BRAF→MEK→ERK pathway activity and tumour-cell proliferation. **b**, **c** and **e** show mean percentage  $\pm$  s.e.m.,  $n = 6$ .

within 10 days after drug withdrawal (Fig. 4b). Consistent with *in vitro* observations, immunoblot analysis of melanoma specimens collected by serial FNAs from each tumour indicated that drug withdrawal led to elevated pMEK1/2→pERK1/2 signalling, concomitant with tumour regression (Fig. 4c and Supplementary Fig. 5a). However, following an initial period of tumour regression after drug withdrawal, tumours showed re-growth—at which time pERK1/2 levels in the vehicle-treated tumours (Fig. 4c and Supplementary Fig. 5a, light blue bars) returned to their original levels (Fig. 4c and Supplementary Fig. 5a, dark blue bars). These data support the hypothesis that vemurafenib-resistant tumours suffer a fitness deficit in the absence of vemurafenib. On the basis of molecular analysis of the HMEX1906 model, we propose that resistance to BRAF inhibitors is due to increased BRAF(V600E) expression. To expand on observations made using the HMEX1906 model, we tested whether this phenomenon might hold true in additional models of vemurafenib resistance. First, we assessed the effect of vemurafenib withdrawal from SK-MEL239-C3 cells, in which resistance is due to expression of a 61-kDa splice variant of BRAF(V600E) (Supplementary Fig. 6b)<sup>8</sup>. In a clonogenicity assay, we observed significantly fewer SK-MEL239-C3 cell colonies when cultured in the absence of vemurafenib (Supplementary Fig. 6c). In addition, we tested the effects of vemurafenib withdrawal from a second BRAF-mutated PDX (M120214) (Supplementary Table 1) isolated from a patient whose melanoma already showed vemurafenib resistance. This PDX was established in mice dosed with 45 mg kg<sup>-1</sup> vemurafenib (twice daily) immediately after tumour implantation. After 59 days of drug treatment, drug administration was ceased in four out of five tumour-bearing mice. All four tumours demonstrated clear signs of drug-withdrawal-induced tumour regression (Supplementary Fig. 6c), consistent with observations in the HMEX1906 model (Fig. 4c). These models support the observation that vemurafenib-resistant 45V-RT melanomas show a fitness deficit in the absence of vemurafenib.

One prediction of this model is that, whereas continuous vemurafenib treatment will inevitably select for drug-resistant tumour cells, discontinuous dosing would create a disadvantageous environment for drug-resistant cells—thereby forestalling the onset of lethal drug resistance. To that end, mice were implanted with parental HMEX1906 tumours and treated either continuously or intermittently (4 weeks on, 2 weeks off) with 15 mg kg<sup>-1</sup> vemurafenib (twice daily), such that mice on the intermittent dosing schedule received the same or a greater cumulative drug dose as mice on the continuous schedule over the entire treatment period. As predicted, mice continuously dosed with vemurafenib developed lethal drug-resistant disease within 100 days after initiation of drug administration. By contrast, none of the mice on the intermittent dosing schedule developed drug-resistant disease over the course of 200 days (Figs 1b and 4d, and Supplementary Fig. 5b). In addition, a similar intermittent versus continuous dosing experiment was conducted in another early passage PDX expressing BRAF(V600E), HMEX2613 (Supplementary Fig. 6a and Supplementary Table 1). In this case, the intermittent dosing schedule was individualized for each tumour-bearing mouse. As in the HMEX1906 model, HMEX2613 tumours treated continuously with 45 mg kg<sup>-1</sup> vemurafenib (twice daily) developed lethal drug-resistant disease, whereas mice dosed intermittently with vemurafenib did not (Fig. 4e). Irrespective of the underlying mechanism of resistance in the two models, these results indicate that intermittent dosing significantly delays the onset of drug resistance by exploiting the fitness deficit shown by drug-resistant tumour cells in the absence of drug. Furthermore, although we observed that counter-selection against resistant cells by cessation of vemurafenib administration allowed drug-sensitive tumours to restart their growth, these cells remained responsive to the antitumour effects of vemurafenib re-administration.

Although vemurafenib can inhibit BRAF(V600E)→MEK→ERK signalling sufficiently to elicit marked tumour regression, the durability



**Figure 4 | Intermittent dosing of vemurafenib can be exploited to forestall the development of drug resistance *in vivo*.**

**a**, Vemurafenib-resistant tumours were implanted and then mice were dosed with either vehicle or 45 mg kg<sup>-1</sup> vemurafenib twice daily (mean percentage  $\pm$  s.e.m.,  $n = 30$ ) and monitored for tumour establishment over a period of 100 days. **b**, Vemurafenib-resistant tumours were implanted into nude mice and dosed with 45 mg kg<sup>-1</sup> vemurafenib twice daily immediately after implant. Once tumours reached a volume of  $\sim 1,500$  mm<sup>3</sup>, mice were switched from vemurafenib to vehicle control (blue line), while one mouse remained on vemurafenib (red line). FNAs (purple arrows) were taken from the tumours before and after drug withdrawal to evaluate pERK. **c**, Lysates collected from the FNA were used to measure pERK, bars represent the pERK1/2 levels from seven different tumours (separated by dotted grey lines), while mice were dosed with

vemurafenib (dark blue bars) or vehicle (light blue bars). The growth kinetics for each tumour is represented by the line graph above the pERK1/2 bars and FNA sampling is depicted by arrows (dark blue, on drug; light blue, off drug). **d**, Tumour growth kinetics of naive parental HMEX1906 tumours with seven tumours dosed continuously (top) and nine tumours dosed intermittently (bottom). Intermittent dosing of vemurafenib was carried out on a 4-week on drug (green arrow) and 2-week off drug (red arrow) schedule with 15 mg kg<sup>-1</sup> vemurafenib twice daily. **e**, Kaplan-Meier curve of data in **d** ( $n = 7$ , continuous dosing and  $n = 9$ , intermittent dosing) and Supplementary Fig. 6a ( $n = 7$ , continuous dosing and  $n = 8$ , intermittent dosing), shows that there is a significant survival advantage with an intermittent dosing (solid lines) compared to a continuous dosing schedule (dashed lines). The end point for euthanasia was predetermined as a tumour size of 1,200 mm<sup>3</sup>.

of vemurafenib responses is limited by acquired drug resistance<sup>6–9,15,16</sup>. Our results suggest that the proliferation of vemurafenib-resistant cells can be dependent on the continuous presence of the drug, such that tumour growth is inhibited after cessation of drug administration. These data are consistent with previous results indicating that both normal and tumour cells can be sensitive to both the quality (that is,

which pathways are activated) and the quantity (that is, magnitude of pathway activation) of signal pathway activation<sup>12–14,17–19</sup>. Furthermore, we show that discontinuous dosing forestalls the onset of drug resistance in two primary human xenograft models. Our observations, and those of others, suggest that the majority of BRAF(V600E) melanomas remain reliant on the reactivation of ERK despite ongoing inhibition of



BRAF(V600E). In these cases, drug resistance is achieved via elevated signalling through receptor tyrosine kinases, mutational activation of *NRAS* or *MEK*, amplification of *BRAF*, or alternative splicing of the *BRAF*<sup>T1799A</sup> precursor mRNA to yield aberrant forms of BRAF(V600E)<sup>6–9,15,16</sup>. Our data indicate that some mechanisms of vemurafenib resistance confer a fitness deficit upon the tumour cells in the absence of the drug. This is probably due to elevated ERK1/2 activation that leads to arrest of the cell division cycle or the onset of apoptosis. Indeed, established literature indicates that deliberate elevation of RAF→MEK→ERK signalling in bona fide human cancer cells can have antiproliferative effects<sup>11–13</sup>. Furthermore, these data suggest that the durability of responses to agents like vemurafenib may be improved through alterations in the dosing schedule, a phenomenon consistent with a recent case report of two melanoma patients with *BRAF* mutations who demonstrated a secondary antitumour response to BRAF inhibition after cessation of BRAF-inhibitor treatment owing to acquired drug resistance<sup>20</sup>. Moreover, these results may have implications for other targeted cancer therapies, especially those that target RAF→MEK→ERK signalling. Although published clinical observations are still lacking, we suggest that dose regimens that exceed the daily maximum tolerated dose could be used to induce rapid tumour regression, followed by a drug holiday to prevent the onset of toxicities observed with chronic daily dosing and the emergence of drug-resistant tumour cells. Whereas continuous dosing promotes the clonal expansion of drug-resistant cells<sup>12–14,21,22</sup>, intermittent dosing could serve to eliminate the fitness advantage of the resistant cells and delay the onset of drug-resistant disease. Hence, our results could have an impact on the use of pathway-targeted therapies to treat at least the subset of melanomas in which *BRAF* is mutated.

## METHODS SUMMARY

**Mouse experiments and drug administration.** All laboratory animal work was conducted under appropriate United States Department of Agriculture (USDA), the Office of Laboratory Animal Welfare (OLAW), the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) (full accreditation since 1998) and The Guide for the Care and Use of Laboratory Animals laws and guidelines. Female nu/nu mice were purchased from Charles River. Mice were administered vemurafenib by oral gavage (PO) twice daily. Copy number assays, RT-qPCR and pERK1/2 measurements of tumours were carried out on DNA, RNA or protein samples collected from individual mouse tumours (Figs 1c and 2b, c, and Supplementary Fig. 2a, b).

**Implanting and harvesting tumour samples.** HMEX1906 is a tumour xenograft model developed from a lymph node metastatic melanoma biopsy sample. Tumour pieces were fragmented into 2-mm<sup>3</sup> chunks and transferred into a drop of BD Matrigel before subcutaneous implantation in the right suprascapular region of female nude mice using a trocar. Tumours were measured with digital callipers twice a week and once they reached 1,000–2,000 mm<sup>3</sup>, the animal was euthanized and the tumour was harvested for further analysis, re-implantation or freeze-back. **Tumour collection for analysis.** Freshly harvested tumours were immediately pulverized over liquid nitrogen. For protein analysis, 100 mg of tumour powder was lysed in RIPA buffer (50 mM Tris at pH 7.4, 100 mM NaCl and 0.1% SDS) supplemented with a complete protease inhibitor mixture (Roche) and phosphatase inhibitor (PhosStop by Roche), and centrifuged at 13,000g for 10 min at 4 °C. The suspension was then homogenized using MagNa Lyser Green Beads (Roche). **FNAs.** Tumour-bearing mice were anaesthetized using isoflurane and samples were collected using a 21-gauge needle. The sample was then immediately flushed into RIPA buffer for protein analysis or RLT buffer for mRNA analysis, or cell culture media to make cell lines.

**Full Methods** and any associated references are available in the online version of the paper.

Received 19 May; accepted 29 November 2012.

Published online 9 January; corrected online 13 February 2013 (see full-text HTML version for details).

1. Fecher, L. A., Amaravadi, R. K. & Flaherty, K. T. The MAPK pathway in melanoma. *Curr. Opin. Oncol.* **20**, 183–189 (2008).

2. Davies, H. *et al.* Mutations of the *BRAF* gene in human cancer. *Nature* **417**, 949–954 (2002).
3. Sosman, J. A. *et al.* Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *N. Engl. J. Med.* **366**, 707–714 (2012).
4. Bollag, G. *et al.* Clinical efficacy of a RAF inhibitor needs broad target blockade in *BRAF*-mutant melanoma. *Nature* **467**, 596–599 (2010).
5. Flaherty, K. T. *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.* **363**, 809–819 (2010).
6. Nazarian, R. *et al.* Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468**, 973–977 (2010).
7. Wagle, N. *et al.* Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J. Clin. Oncol.* **29**, 3085–3096 (2011).
8. Poulidakos, P. I. *et al.* RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). *Nature* **480**, 387–390 (2011).
9. Shi, H. *et al.* Melanoma whole-exome sequencing identifies *V600E*-BRAF amplification-mediated acquired B-RAF inhibitor resistance. *Nature Commun.* **3**, 724 (2012).
10. Bennett, D. C. How to make a melanoma: what do we know of the primary clonal events? *Pigment Cell Melanoma Res.* **21**, 27–38 (2008).
11. Petti, C. *et al.* Coexpression of NRAS<sup>Q61R</sup> and BRAF<sup>V600E</sup> in human melanoma cells activates senescence and increases susceptibility to cell-mediated cytotoxicity. *Cancer Res.* **66**, 6503–6511 (2006).
12. Zhu, J., Woods, D., McMahon, M. & Bishop, J. M. Senescence of human fibroblasts induced by oncogenic Raf. *Genes Dev.* **12**, 2997–3007 (1998).
13. Woods, D. *et al.* Raf-induced proliferation or cell cycle arrest is determined by the level of Raf activity with arrest mediated by p21Cip1. *Mol. Cell. Biol.* **17**, 5598–5611 (1997).
14. Marshall, C. J. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal regulated kinase activation. *Cell* **80**, 179–185 (1995).
15. Johannessen, C. M. *et al.* COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature* **468**, 968–972 (2010).
16. Villanueva, J. *et al.* Acquired resistance to BRAF inhibitors mediated by a RAF kinase switch in melanoma can be overcome by cotargeting MEK and IGF-1R/PI3K. *Cancer Cell* **18**, 683–695 (2010).
17. Tap, W. D. *et al.* Pharmacodynamic characterization of the efficacy signals due to selective BRAF inhibition with PLX4032 in malignant melanoma. *Neoplasia* **12**, 637–649 (2010).
18. Hoeflich, K. P. *et al.* Oncogenic BRAF is required for tumor growth and maintenance in melanoma models. *Cancer Res.* **66**, 999–1006 (2006).
19. Hingorani, S. R., Jacobetz, M. A., Robertson, G. P., Herlyn, M. & Tuveson, D. A. Suppression of BRAF(V599E) in human melanoma abrogates transformation. *Cancer Res.* **63**, 5198–5202 (2003).
20. Neyns, B., Seghers, A. C., Wilgenhof, S. & Lebbe, C. Successful rechallenge in two patients with BRAF-V600-mutant melanoma who experienced previous progression during treatment with a selective BRAF inhibitor. *Melanoma Res.* **22**, 466–472 (2012).
21. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
22. Chmielecki, J. *et al.* EGFR-mutant lung adenocarcinomas treated first-line with the novel EGFR inhibitor, XL647, can subsequently retain moderate sensitivity to erlotinib. *J. Thorac. Oncol.* **7**, 434–442 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the members of the Novartis Institutes for BioMedical Research (NIBR) Pharmacology department for technical support, comments and discussions during the course of this work. We thank C. Voliva, N. Aziz and E. Collinson for discussions. We thank B. Weisburd and the rest of the NIBR Bioinformatics department for assistance with exome sequencing data analysis. We thank S. Kaufman for sharing her knowledge of cell-based assays. We thank V. Marsh, N. Rosen, P. Poulidakos and D. Solit for providing additional advice and reagents. M.D.T. was supported by an NIBR Presidential Postdoctoral Fellowship. M.M. acknowledges support from the Melanoma Research Alliance and the National Cancer Institute (R01-CA176839). A.S.L. was supported by a National Research Service Award T32 training grant HL007185.

**Author Contributions** M.D.T., M.M. and D.D.S. designed all experiments. M.D.T. performed *in vivo* and *in vitro* experiments and collected data. M.D.T., M.M. and D.D.S. analysed data, wrote the paper and guided the manuscript through review. F.S. assisted in performing *in vivo* experiments. A.S.L. carried out the clonogenic assay with the SK-Mel-239-C3 cells. M.P.L. and R.D. provided the human patient biopsy samples, and R.D. assisted with data analysis and interpretation. W.R.S. and N.K.P. provided input on the experimental approach and on the manuscript. M.M. and D.D.S. are co-senior authors of this manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.D.S. ([darrin.stuart@novartis.com](mailto:darrin.stuart@novartis.com)) or M.M. ([mcmahon@cc.ucsf.edu](mailto:mcmahon@cc.ucsf.edu)).

## METHODS

**Copy number assay.** DNA from tumour tissue was extracted using the DNeasy Blood and Tissue Kit (69504, Qiagen) according to the manufacturer's protocol. Quantitative PCR was carried out using the Taqman genotyping master mix, Taqman assay (Hs04949885\_cn, Hs05005955\_cn or Hs04949201\_cn; Applied Biosystems), and RnaseP was used as a normalization control for DNA content. Quantitative analysis was carried out using the 7500 Real-time PCR system (Applied Biosystems).

**RT-qPCR.** RNA was extracted using the RNeasy Kit (74104, Qiagen) according to the manufacturer's protocol. One-step RT-qPCR reactions were carried out in triplicate using the Quantitect Multiple RT-PCR master mix, Taqman Gene expression assay primer and 18S probe, Taqman Gene expression assay primer and BRAF probe (4331182, Applied Biosystems), and Quantitect RT mix. Quantitative measurements were collected using the 7500 Realtime PCR system (Applied Biosystems). Endogenous control 18S was used as a normalization control for RNA content.

**pERK1/2 and pMEK measurements.** Meso Scale Discovery plates were used for pERK1/2 (K111DWD-2), total ERK (K111DXX-2), pMEK (K111DUD-2) or total MEK (K111CWD-2) analysis according to the manufacturer's protocol. Plates were analysed on the SECTOR Imager. Both pERK1/2 and pMEK readings were normalized to the total ERK and total MEK levels, respectively. Data for Supplementary Figs 1d, e, 2 and 4a, c were collected using this method.

**Cell viability assay.** HMEX1906 cells were split 1:2 the day before seeding. Cells were plated in 100 µl of media the next day at 2,000 cells per well onto black-walled, clear-bottom 96-well plates (3904, Corning Costar). Cells were incubated for 3 days with or without compound at 37 °C before carrying out the viability assay. Using the Cell Titer-Glo Luminescent Cell Viability assay kit (Promega) and instructions, luminescent measurements were taken on Trilux MicroBeta2. The graphically represented values are means  $\pm$  s.d. for three independent samples.

**Cell culture.** HMEX1906 cell lines were generated using FNAs from either the parental or the resistant tumours. The FNA was then directly flushed out into EGM media (CC-3124, Lonza) and transferred onto collagen-coated plates (BD Biosciences). Media was changed every day until all tumour debris was gone. Once

cell lines were established the plates were maintained at about 50–80% confluence and with media change twice a week. Resistant tumour cell lines were maintained in 50 nM vemurafenib.

**BRAF siRNA and western blots.** HMEX1906 cells were plated 1 day before transfection at 70% confluency. The next day parallel plates were left untreated, or were transfected with a non-targeting pool of siRNA or BRAF14 on target plus siRNA (target sequence, AGACGGGACUCGAGUGAUG, J-003460-14; Dharmacon). For the transfections, 1,000 µl of Opti-MEM was mixed with 17.5 µl of siRNA to give a final concentration of 50 nM; this was then combined with 1,000 µl Opti-MEM and 21 µl of Dharmafect 1. After a 20-min incubation at room temperature (25 °C) on a shaker, the transfection mix was applied drop-wise to the cells and incubated overnight. Protein lysates, pERK1/2 assays and cell viability assays were then carried out 72 h after transfection. BRAF immunoblot was carried out with RAF-B (F-7) antibody (sc-5284; Santa Cruz).

**Accumen pERK1/2 assay.** HMEX1906 cells were plated at 2,000 cells per well onto black-walled, clear-bottom Corning Costar 96-well plates (#3904) and incubated for 72 h in varying drug concentrations. On day 3, the media was discarded and the cells were fixed using 100 ml of 4% paraformaldehyde in PBS for 15–20 min. The wells were washed with PBS and then permeabilized with PBS plus 0.1% Triton for 10–15 min at room temperature. The plates were then blocked with 5% normal goat serum for 1 h, after which pERK1/2 (4370; Santa Cruz) primary antibody was applied to the cells at a 1:200 dilution in PBS with 0.1% Triton and 1% BSA. Plates were left overnight on a shaker at 4 °C. Plates were washed and secondary antibody (Invitrogen Alexa 488) and Hoechst stain (34580; Invitrogen) were applied at 1:1,000 and 1:2,500, respectively, for 1–1.5 h. The plates were washed and sealed to scan on the Accumen EX3. pERK1/2 levels were normalized to cell numbers for data analysis.

**Clonogenic assay.** SK-Mel-239-C3 vemurafenib-resistant cells were plated at the indicated cell density in 2 µM vemurafenib. The next day, half the plates were washed and re-fed with media lacking vemurafenib (day 0). Plates were stained with crystal violet on the days indicated. Plates were re-fed with the appropriate media plus or minus vemurafenib every 3 days.

# Biguanides suppress hepatic glucagon signalling by decreasing production of cyclic AMP

Russell A. Miller<sup>1</sup>, Qingwei Chu<sup>1</sup>, Jianxin Xie<sup>2</sup>, Marc Foretz<sup>3,4,5</sup>, Benoit Viollet<sup>3,4,5</sup> & Morris J. Birnbaum<sup>1</sup>

Glucose production by the liver is essential for providing a substrate for the brain during fasting. The inability of insulin to suppress hepatic glucose output is a major aetiological factor in the hyperglycaemia of type-2 diabetes mellitus and other diseases of insulin resistance<sup>1,2</sup>. For fifty years, one of the few classes of therapeutics effective in reducing glucose production has been the biguanides, which include phenformin and metformin, the latter the most frequently prescribed drug for type-2 diabetes<sup>3</sup>. Nonetheless, the mechanism of action of biguanides remains imperfectly understood. The suggestion a decade ago that metformin reduces glucose synthesis through activation of the enzyme AMP-activated protein kinase (AMPK) has recently been challenged by genetic loss-of-function experiments<sup>4</sup>. Here we provide a novel mechanism by which metformin antagonizes the action of glucagon, thus reducing fasting glucose levels. In mouse hepatocytes, metformin leads to the accumulation of AMP and related nucleotides, which inhibit adenylate cyclase, reduce levels of cyclic AMP and protein kinase A (PKA) activity, abrogate phosphorylation of critical protein targets of PKA, and block glucagon-dependent glucose output from hepatocytes. These data support a mechanism of action for metformin involving antagonism of glucagon, and suggest an approach for the development of antidiabetic drugs.

Biguanides exert their major effect through inhibition of liver glucose production, although enhanced glucose disposal has also been noted in some studies<sup>5,6</sup>. Despite the wide acceptance of metformin as a first-line therapeutic for diabetes, its mechanism of action remains unclear. Metformin inhibits mitochondrial respiratory complex 1, thus reducing hepatic energy charge; ten years ago it was suggested that metformin functioned through activation of the kinase AMPK<sup>7–10</sup>. This has been challenged recently in experiments using livers and primary hepatocytes lacking either AMPK or its upstream activating enzyme LKB1 (ref. 4). Because both glycogenolysis and gluconeogenesis are controlled during the fasting state in part by the hormone glucagon, the abnormal secretion of which is a major factor in the pathophysiology of hyperglycaemia in type-2 diabetes, we explored the idea that metformin produces its effects by inhibiting glucagon signalling pathways<sup>11–13</sup>.

Binding of glucagon to its receptor on the hepatocyte plasma membrane leads to activation of adenylyl cyclase, production of the second-messenger cyclic AMP (cAMP) and stimulation of PKA, which phosphorylates protein targets that work in concert to increase hepatic glucose output<sup>12</sup>. In primary mouse hepatocytes, we tested a panel of compounds that activate AMPK by decreasing energy charge or mimicking a reduced energy charge for the inhibition of glucagon-dependent increases in cAMP. Treatment with all agents increased AMPK phosphorylation and antagonized the glucagon-dependent elevation in cAMP (Supplementary Fig. 1a, b). Phenformin showed a dose-dependent inhibition of glucagon-induced cAMP accumulation with a half-maximal inhibitory concentration of ~150  $\mu\text{M}$ , which correlated well with its effect on intracellular AMP (Fig. 1a, b). To mimic more closely the chronic treatment of diabetic patients with biguanides,

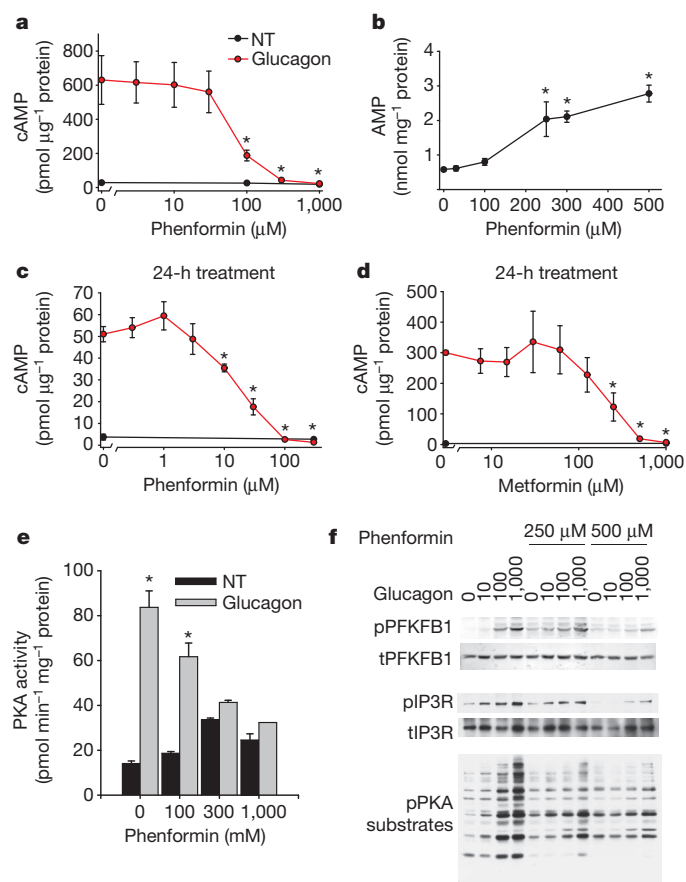
we tested 24-h exposure of hepatocytes to phenformin or metformin for a reduction in glucagon-increased cAMP levels. After long-term treatment, phenformin was effective at considerably lower concentrations, with concentrations of 10  $\mu\text{M}$  or greater causing significant reductions in glucagon-stimulated increases in cAMP levels (Fig. 1c). Metformin also inhibited cAMP accumulation at concentrations of 125  $\mu\text{M}$  or greater, levels slightly higher than in the serum of diabetic patients after taking a 1 mg dose of metformin but probably similar to those accumulated in splanchnic tissues<sup>14,15</sup> (Fig. 1d). In rats, administration of a therapeutic dose of 50 mg kg<sup>-1</sup> leads to levels greater than 250  $\mu\text{M}$  in the liver<sup>14,15</sup>. Like phenformin, therapeutic concentrations of metformin elicited significant increases in AMP levels in primary hepatocytes (Supplementary Fig. 2a–c). Glucagon did not alter adenine nucleotide levels or activate AMPK in primary hepatocytes and did not affect the changes produced by metformin (Supplementary Fig. 2a–d).

To determine the effects of phenformin on the kinetics of activation of PKA, we used the AKAR3 fluorescence resonance energy transfer (FRET) reporter<sup>16</sup>. Primary hepatocytes were infected with adenovirus encoding AKAR3 and 18 h later confocal images were acquired over time (Supplementary Fig. 3a). Glucagon increased the FRET ratio (FRET/cyan fluorescent protein (CFP)) of AKAR3 in the cytoplasm of hepatocytes within 1 min and reached a maximum at 2 min, thereafter exhibiting only minimal decay for 15 min. Phenformin both delayed the rise in PKA activity and accelerated its decay (Supplementary Fig. 3b). The concentration dependence of inhibition of PKA activity measured biochemically was similar to blockade of cAMP accumulation in primary hepatocytes (Fig. 1e). AICAR, an adenosine analogue that can be phosphorylated to form the AMP mimetic ZMP, also antagonized the glucagon-dependent increase of PKA activity in isolated hepatocytes (Supplementary Fig. 4a).

Treatment of primary hepatocytes with phenformin inhibited the glucagon-dependent phosphorylation of cellular proteins at PKA substrate motifs (Fig. 1f). Phenformin also antagonized phosphorylation of the PKA substrates PFKFB1 and the inositol-1,4,5-trisphosphate receptor IP3R (also known as ITPR1), as revealed by western blots using phospho-specific antibodies against these proteins (Fig. 1f and Supplementary Fig. 5a). If reductions in levels of endogenous cAMP are important for the actions of biguanides, we would expect the latter to block PKA target phosphorylation in response to glucagon but not a membrane-permeable analogue of cAMP, SP-8Br-cAMPS-AM. Glucagon induced phosphorylation of the PKA target proteins PFKFB1, CREB (also known as CREB1) and IP3R in primary hepatocytes; phenformin activated AMPK and inhibited phosphorylation of these PKA target proteins (Fig. 2a and Supplementary Fig. 6a). In contrast, when SP-8Br-cAMPS-AM concentrations were used that elicited comparable phosphorylation of the PKA substrates, the actions of SP-8Br-cAMPS-AM were not antagonized by phenformin (Fig. 2b and Supplementary Fig. 6b). These results demonstrate that phenformin inhibits the glucagon- but not SP-8Br-cAMPS-AM-dependent phosphorylation of PKA substrates, a result most consistent with biguanides working at a signalling step upstream from the activation

<sup>1</sup>Institute for Diabetes, Obesity, and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>2</sup>Cell Signaling Technology, Inc., 3 Trask Lane, Danvers, Massachusetts 01923, USA. <sup>3</sup>Inserm, U1016, Institut Cochin, Paris 75014, France. <sup>4</sup>Cnrs, UMR8104, Paris 75014, France. <sup>5</sup>Université Paris Descartes, Sorbonne Paris cité, Paris 75006, France.

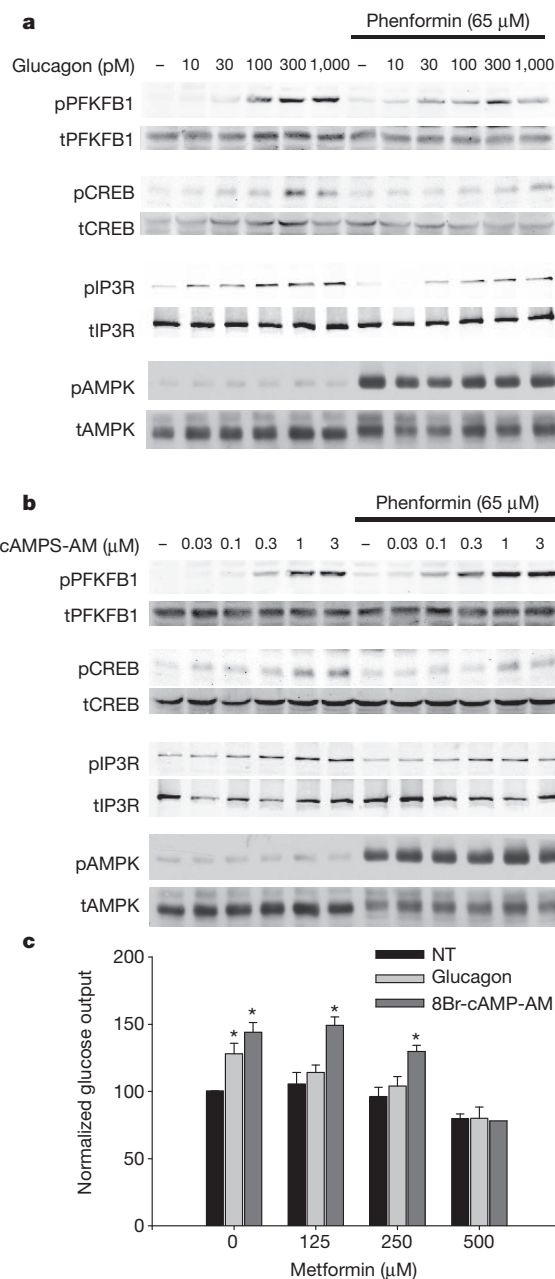




**Figure 1 | Biguanides inhibit cAMP accumulation.** **a**, Primary hepatocytes were incubated with the indicated phenformin concentrations for 2 h, 5 nM glucagon or no treatment (NT) for 15 min, lysed, and assayed for total cellular cAMP and protein. **b**, Primary hepatocytes incubated with the indicated concentration of phenformin for 2 h were extracted with perchloric acid and cellular nucleotides quantified by high-performance liquid chromatography (HPLC). **c, d**, Primary hepatocytes were incubated with the indicated concentration of phenformin (c) or metformin (d) for 24 h, treated with 5 nM glucagon, lysed, and assayed for total cellular cAMP. **e**, Primary hepatocytes were incubated with the indicated concentrations of phenformin for 2 h, treated with 5 nM glucagon, lysed, and PKA kinase activity determined. **f**, Primary hepatocytes were incubated with phenformin for 2 h, then with glucagon, and protein was analysed by western blot with the phospho- (p)PKA substrate motif antibody, total (t) and phospho- (p)PFKFB1 antibodies, and total and phospho-IP3R antibodies. Error bars represent standard error of the mean (s.e.m.).

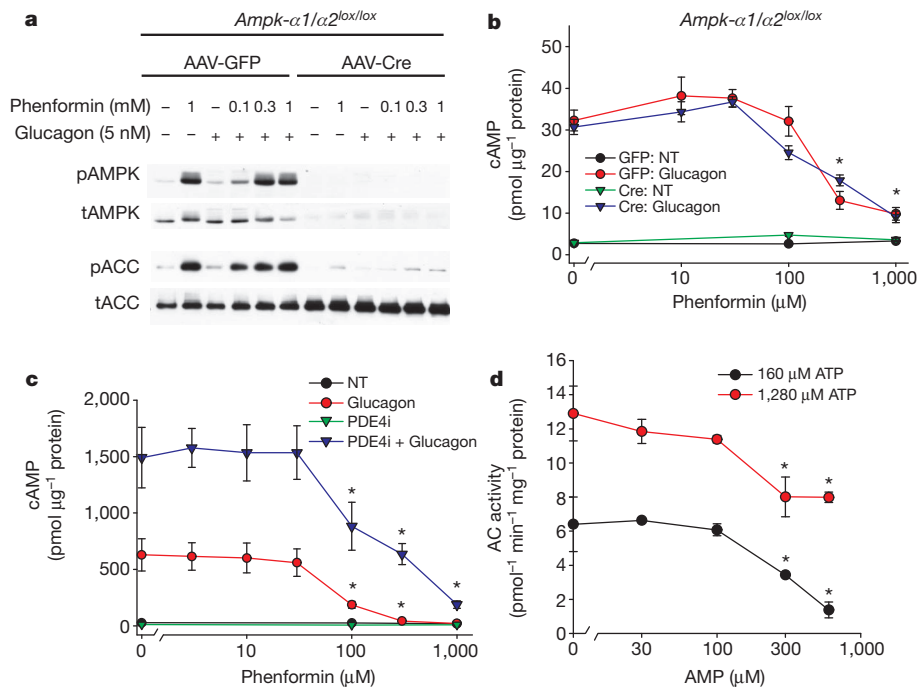
of PKA. Both glucagon and SP-8Br-cAMPS-AM increased glucose output from primary hepatocytes; however, treatment of hepatocytes with therapeutic concentrations of metformin prevented glucagon- but not SP-8Br-cAMPS-AM-stimulated increases in glucose output (Fig. 2b). Of note, higher concentrations of metformin reduced basal as well as glucagon- and SP-8Br-cAMPS-AM-dependent glucose output. If the major actions of biguanide were mediated by decreases in cAMP, one would expect the drug to be ineffective in cells in which PKA has already been inhibited. In hepatocytes overexpressing a dominant-negative PKA regulatory subunit that is unable to bind cAMP, PKA activity was necessary for glucagon stimulation of glucose output, and biguanides did not reduce the rate of glucose output any further (Supplementary Fig. 6c).

As the biguanides and other drugs we used activated AMPK in parallel to the reduction in cAMP, we asked whether the kinase was required for the effects of biguanides. Mice homozygous for the floxed



**Figure 2 | Biguanides inhibit glucagon signalling.** **a, b**, Primary hepatocytes were cultured for 18 h in the presence or absence of 65  $\mu$ M phenformin and for 15 min with the indicated concentrations of glucagon (a) or the cell-permeable PKA agonist SP-8Br-cAMPS-AM (b). Western blot analysis of total (t) and phosphorylated (p) PFKFB1, CREB, IP3R and AMPK. **c**, Cells were treated with the indicated concentration of metformin and either 1 nM glucagon or 3  $\mu$ M SP-8Br-cAMPS-AM, or were left untreated (NT), for 14 h and then glucose output measured for 5 h. Data represent the means of three experiments,  $N = 6$  for each experiment. Error bars represent s.e.m.

alleles of both catalytic  $\alpha 1$  and  $\alpha 2$  subunits of the AMPK complex were infected with adeno-associated virus expressing Cre recombinase, and western blots confirmed deletion of AMPK  $\alpha$  protein and loss of phenformin-dependent phosphorylation of the AMPK substrate acetyl-CoA carboxylase (ACC) (Fig. 3a). In hepatocytes lacking any detectable AMPK activity, phenformin blocked glucagon-dependent cAMP accumulation in a manner indistinguishable from that in control cells (Fig. 3b). AICAR also reduced cAMP in the absence of AMPK (Supplementary Fig. 4b). These data show that the effects of biguanides on cAMP metabolism are independent of AMPK.



**Figure 3 | Mechanism of biguanide effect on cAMP production.** **a**, *Ampk-α1/α2<sup>lox/lox</sup>* mice were infected with AAV-TBG-GFP or AAV-TBG-Cre virus and 14 days later primary hepatocytes were isolated. Cells were treated with the indicated concentrations of phenformin for 2 h followed by 5 nM glucagon or no treatment (NT) for 15 min. **a**, Total cellular protein was analysed by western blot for total (t) and phosphorylated (p) T172 AMPK and total and phospho-S79 ACC. **b**, Hepatocytes were lysed and total cellular cAMP levels were quantified by ELISA. *N* = 4 for all points. **c**, Primary hepatocytes were

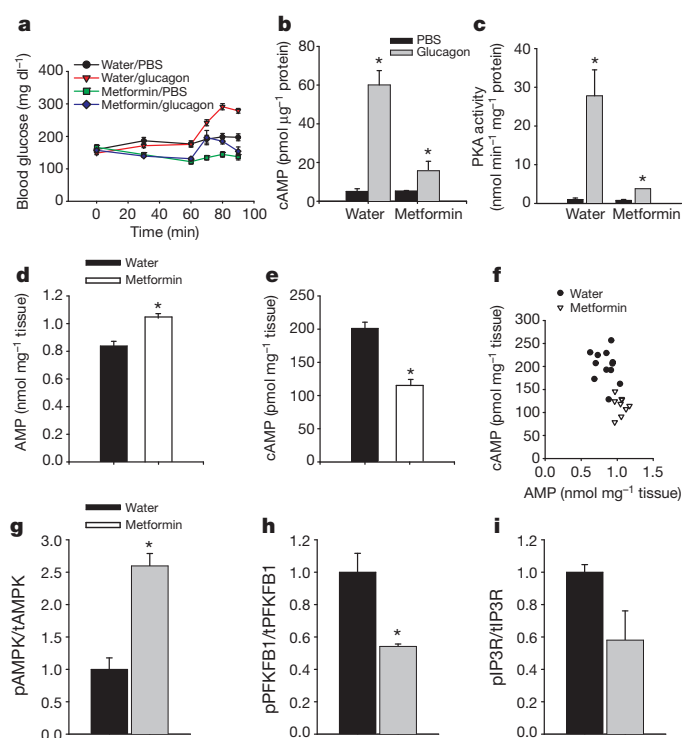
incubated with the indicated concentrations of phenformin for 2 h and 50 μM RO-20-1724 (PDE4 inhibitor; PDE4i) for the final 30 min. Cells were then treated with 5 nM glucagon for 15 min, lysed, and total cellular cAMP was assayed. *N* = 4 for all points. **d**, The membrane fraction of primary hepatocytes was isolated by differential centrifugation and used in assays for adenylyl cyclase activity in the presence of the indicated AMP and ATP concentrations, 100 nM glucagon and 100 μM GTP. *N* = 6 for all points. Error bars represent s.e.m.

To determine whether reduction of the glucagon-dependent increase in cAMP was caused by activation of cAMP phosphodiesterase (PDE), we used pharmacological inhibitors of PDEs. If inhibition of cAMP degradation did not also block the effect of phenformin on cAMP, it would imply that the site of action was unlikely to be PDE. Exposure of glucagon-treated hepatocytes to 3-isobutyl-1-methylxanthine (IBMX; a non-specific PDE inhibitor) or Ro-20-1724 (a specific PDE4 class inhibitor), but not cilostamide (a specific PDE3 class inhibitor), increased cAMP levels, indicating that PDE4 is the major cAMP-degrading enzyme in these cells (Supplementary Fig. 7a). Even in the presence of IBMX or Ro-20-1724, phenformin inhibited glucagon-stimulated accumulation of cAMP (Supplementary Fig. 7a). Moreover, Ro-20-1724 did not affect the dose responsiveness of hepatocytes to phenformin (Fig. 3c). To resolve phenformin's site of action further, we measured its effects on cAMP accumulation in response to forskolin, which binds to and stimulates adenylyl cyclase independent of the activated G $\alpha$  protein. Both phenformin and AICAR blocked cAMP accumulation by both glucagon and forskolin to the same extent, indicating that the effects of these drugs were unlikely to be due to modulation of the glucagon receptor or G-protein signalling (Supplementary Figs 4c and 7b).

As phenformin blocked glucagon-induced cAMP production at concentrations that correlated well with its effect on AMP levels, we sought to determine whether AMP could directly inhibit adenylyl cyclase, as has been reported previously<sup>17–20</sup>. Forskolin and glucagon stimulated adenylyl cyclase activity about fourfold in membranes isolated from primary mouse hepatocytes (Supplementary Fig. 8a). AMP inhibited glucagon-stimulated adenylyl cyclase activity when ATP was present at 160 μM, which represents typical conditions for the assay, as well as at the more physiological concentration of 1.28 mM ATP (Fig. 3d). To determine whether AMP accumulated in the livers of metformin-treated mice to levels sufficient to inhibit

adenylyl cyclase, we measured hepatic AMP in fasted wild-type mice. Consistent with a recent report, we found the concentration of AMP to be 2.3 mM and to rise to 2.9 mM after treatment with metformin in mice fed a normal chow diet<sup>21,22</sup> (Supplementary Table 1). Because these values seemed unphysiologically high, we also measured nucleotides in primary hepatocytes (Supplementary Table 2). In untreated cells, the AMP concentration was 215 μM, which correlated well with a value of 300 μM reported previously<sup>23</sup> in perfused rat liver using <sup>31</sup>P NMR. The AMP concentration in hepatocytes rose to over 1 mM in a dose-dependent manner upon treatment with phenformin (Supplementary Table 2).

To test whether the effects of biguanides on the accumulation of cAMP in primary hepatocytes were representative of its actions *in vivo*, we examined the effects of metformin after injection of glucagon into mice. Glucagon rapidly increased blood glucose, an effect that was blunted by pre-administration of metformin (Fig. 4a). Gavage of mice with metformin increased both AMPK and ACC phosphorylation, indicating a decrease in the cellular energy charge of the liver (Supplementary Fig. 9a). Most importantly, metformin abrogated the subsequent glucagon-dependent increase in hepatic cAMP, PKA activity and phosphorylation of PKA substrates, demonstrating *in vivo* blockade of this pathway by biguanides (Fig. 4b, c). To assess the effect of metformin on the action of endogenous glucagon, fasted mice were given an oral gavage of water or 250 mg kg<sup>-1</sup> metformin, and 1 h later their blood glucose levels were determined and hepatic tissue was collected. This dose of metformin caused a significant drop in the blood glucose levels and an elevation in hepatic AMP (Supplementary Fig. 9b and Supplementary Table 1), which correlated well with a concomitant reduction in hepatic cAMP content (Fig. 4d–f). We next tested the effects of metformin in diabetic mice fed a high-fat diet (HFD) for 10 weeks. These mice had elevated fasting glucose levels and hepatic AMP compared to chow-fed animals (Supplementary



**Figure 4 | Biguanides antagonize glucagon signalling *in vivo*.** **a**, Mice were gavaged with 500 mg kg<sup>-1</sup> metformin and 1 h later were injected intraperitoneally with 200 μg kg<sup>-1</sup> glucagon, and glucose levels were measured at the indicated times. *N* = 6 for water/PBS and metformin/glucagon, *N* = 7 for water/glucagon and metformin/PBS. **b**, **c**, Fed mice were fasted for 1 h and gavaged with water or 500 mg kg<sup>-1</sup> body weight of metformin. One-hour later mice were injected intraperitoneally with 2 mg kg<sup>-1</sup> glucagon, and liver tissue was collected 5 min later. Liver was analysed for total hepatic cAMP by ELISA (**b**; *N* = 3 for each group) and total hepatic PKA activity (**c**; *N* = 7, 8, 6 and 7 for water/PBS, water/glucagon, metformin/PBS and metformin/glucagon, respectively). \**P* < 0.05 compared to PBS. **d**–**f**, 18-h fasted mice were gavaged with water or 250 mg kg<sup>-1</sup> metformin, 1 h later liver tissue was collected, hepatic metabolites were extracted with perchloric acid and total hepatic AMP (**d**) and cAMP (**e**) levels were assayed. *N* = 12 and 9 for the water and metformin groups, respectively. **g**–**i**, Mice fed HFD for 10 weeks were fasted overnight, gavaged with either water or 250 mg kg<sup>-1</sup> metformin, and after 1 h liver tissue was collected for western blot analysis of the phosphorylation status of AMPK (**g**), PFKFB1 (**h**) and IP3R (**i**). *N* = 3 for each group. Error bars represent s.e.m.

Table 1 and Supplementary Fig. 10a, b). Fasted HFD mice were given an oral gavage of water or 250 mg kg<sup>-1</sup> metformin and 1 h later blood glucose was determined and hepatic tissue collected. Metformin significantly lowered blood glucose levels and elevated hepatic AMP (Supplementary Fig. 10c and Supplementary Table 1). Importantly, treatment of these mice with metformin led to a reduction in the phosphorylation of two key PKA target proteins, PFKFB1 and the IP3R, and increased AMPK phosphorylation (Fig. 4g–i and Supplementary Fig. 10d). The lack of change in Akt phosphorylation indicated that metformin was not affecting insulin responsiveness acutely (Supplementary Fig. 10d, e).

Understanding the mechanism by which the drug metformin reduces hepatic glucose output is of considerable importance. Using a variety of biochemical and cell biological methods, we have confirmed earlier reports of an effect of biguanides on cAMP<sup>24–27</sup>. Furthermore, we have shown that it is through an elevation in intracellular AMP that metformin substantially abrogates the activation of adenylyl cyclase by glucagon. This results in a reduction in the phosphorylation of key substrates for maintaining hepatic glucose output. For nearly 40 years it has been recognized that molecules containing an adenine moiety bind to the ‘P site’ of adenylyl cyclase and inhibit

its activity<sup>17</sup>. Although the endogenous P-site ligand has been suggested to be AMP, the physiological or pharmacological relevance of this regulatory event has not previously been recognized<sup>17–20,28</sup>. We now provide support for the idea that therapeutic levels of metformin induce a mild energetic stress in hepatocytes, resulting in an increase in AMP concentration to levels capable of directly inhibiting adenylyl cyclase. These studies suggest that the P site of adenylyl cyclase might represent a novel target for the development of therapeutics for the treatment of insulin resistance and type-2 diabetes.

## METHODS SUMMARY

Primary hepatocytes were isolated by collagenase perfusion as described previously<sup>29</sup>. Adenine nucleotides were extracted from cells and liver with perchloric acid and measured by ion-pair reversed-phase (RP)-HPLC. cAMP in primary hepatocytes and frozen liver tissue was measured by ELISA (GE Healthcare) using the manufacturer's lysis buffer. PKA activity was assayed in cell lysates as PKI-sensitive Kemptide phosphorylation. PKA FRET-activity probes were used to examine intracellular PKA activity on a spinning-disc confocal microscope<sup>16</sup>. Adenylyl cyclase assays were performed using adenosine-5'-triphosphate [α-<sup>32</sup>P] (American Radiolabelled Chemicals), and cAMP was quantified as previously described<sup>30</sup>. Glucose output studies in primary hepatocytes from fasted mice were carried out in Krebs buffer containing gluconeogenic substrates (20 mM lactate, 2 mM pyruvate, 10 mM glutamine) and were quantified using hexokinase-based glucose assays (Sigma). For *in vivo* experiments, metformin was gavaged at the indicated dosage and glucagon was injected intraperitoneally at the indicated dosages. Tissues were collected rapidly from anaesthetized mice and frozen in pre-cooled metal clamps. All results are expressed as the mean ± s.e.m. All two-group comparisons were deemed statistically significant by unpaired two-tailed Student's *t*-test if *P* < 0.05.

**Full Methods** and any associated references are available in the online version of the paper.

Received 25 January; accepted 22 November 2012.

Published online 6 January 2013.

- DeFronzo, R. A., Simonson, D. & Ferrannini, E. Hepatic and peripheral insulin resistance: a common feature of type 2 (non-insulin-dependent) and type 1 (insulin-dependent) diabetes mellitus. *Diabetologia* **23**, 313–319 (1982).
- Postic, C., Dentin, R. & Girard, J. Role of the liver in the control of carbohydrate and lipid homeostasis. *Diabetes Metab.* **30**, 398–408 (2004).
- Nathan, D. M. *et al.* Medical management of hyperglycemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy: a consensus statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* **32**, 193–203 (2009).
- Foretz, M. *et al.* Metformin inhibits hepatic gluconeogenesis in mice independently of the LKB1/AMPK pathway via a decrease in hepatic energy state. *J. Clin. Invest.* **120**, 2355–2369 (2010).
- Inzucchi, S. E. *et al.* Efficacy and metabolic effects of metformin and troglitazone in type II diabetes mellitus. *N. Engl. J. Med.* **338**, 867–873 (1998).
- Goodarzi, M. O. & Bryer-Ash, M. Metformin revisited: re-evaluation of its properties and role in the pharmacopoeia of modern antidiabetic agents. *Diabetes Obes. Metab.* **7**, 654–665 (2005).
- El-Mir, M. Y. *et al.* Dimethylbiguanide inhibits cell respiration via an indirect effect targeted on the respiratory chain complex I. *J. Biol. Chem.* **275**, 223–228 (2000).
- Owen, M. R., Doran, E. & Halestrap, A. P. Evidence that metformin exerts its anti-diabetic effects through inhibition of complex I of the mitochondrial respiratory chain. *Biochem. J.* **348**, 607–614 (2000).
- Shaw, R. J. *et al.* The kinase LKB1 mediates glucose homeostasis in liver and therapeutic effects of metformin. *Science* **310**, 1642–1646 (2005).
- Zhou, G. *et al.* Role of AMP-activated protein kinase in mechanism of metformin action. *J. Clin. Invest.* **108**, 1167–1174 (2001).
- D'Alessio, D. The role of dysregulated glucagon secretion in type 2 diabetes. *Diabetes Obes. Metab.* **13** (suppl. 1), 126–132 (2011).
- Jiang, G. & Zhang, B. B. Glucagon and regulation of glucose metabolism. *Am. J. Physiol. Endocrinol. Metab.* **284**, E671–E678 (2003).
- Unger, R. H. & Cherrington, A. D. Glucagonocentric restructuring of diabetes: a pathophysiologic and therapeutic makeover. *J. Clin. Invest.* **122**, 4–12 (2012).
- Tucker, G. T. *et al.* Metformin kinetics in healthy subjects and in patients with diabetes mellitus. *Br. J. Clin. Pharmacol.* **12**, 235–246 (1981).
- Wilcock, C. & Bailey, C. J. Accumulation of metformin by tissues of the normal and diabetic mouse. *Xenobiotica* **24**, 49–57 (1994).
- Allen, M. D. & Zhang, J. Subcellular dynamics of protein kinase A activity visualized by FRET-based reporters. *Biochem. Biophys. Res. Commun.* **348**, 716–721 (2006).
- Fain, J. N., Pointer, R. H. & Ward, W. F. Effects of adenosine nucleosides on adenylyl cyclase, phosphodiesterase, cyclic adenosine monophosphate accumulation, and lipolysis in fat cells. *J. Biol. Chem.* **247**, 6866–6872 (1972).



18. Blume, A. J. & Foster, C. J. Mouse neuroblastoma adenylate cyclase. Adenosine and adenosine analogues as potent effectors of adenylate cyclase activity. *J. Biol. Chem.* **250**, 5003–5008 (1975).
19. Londos, C. & Preston, M. S. Regulation by glucagon and divalent cations of inhibition of hepatic adenylate cyclase by adenosine. *J. Biol. Chem.* **252**, 5951–5956 (1977).
20. Johnson, R. A., Yeung, S. M., Stubner, D., Bushfield, M. & Shoshani, I. Cation and structural requirements for P site-mediated inhibition of adenylate cyclase. *Mol. Pharmacol.* **35**, 681–688 (1989).
21. Berglund, E. D. *et al.* Hepatic energy state is regulated by glucagon receptor signaling in mice. *J. Clin. Invest.* **119**, 2412–2422 (2009).
22. Stoll, B., Gerok, W., Lang, F. & Haussinger, D. Liver cell volume and protein synthesis. *Biochem. J.* **287**, 217–222 (1992).
23. Masson, S. & Quistorff, B. The  $^{31}\text{P}$  NMR visibility of ATP in perfused rat liver remains about 90%, unaffected by changes of metabolic state. *Biochemistry* **31**, 7488–7493 (1992).
24. Gawler, D. J., Wilson, A. & Houslay, M. D. Metformin treatment of lean and obese Zucker rats modulates the ability of glucagon and insulin to regulate hepatocyte adenylate cyclase activity. *J. Endocrinol.* **122**, 207–212 (1989).
25. Torres, T. P. *et al.* Impact of a glycogen phosphorylase inhibitor and metformin on basal and glucagon-stimulated hepatic glucose flux in conscious dogs. *J. Pharmacol. Exp. Ther.* **337**, 610–620 (2011).
26. Yu, B., Pugazhenth, S. & Khandelwal, R. L. Effects of metformin on glucose and glucagon regulated gluconeogenesis in cultured normal and diabetic hepatocytes. *Biochem. Pharmacol.* **48**, 949–954 (1994).
27. Zhang, T. *et al.* Mechanisms of metformin inhibiting lipolytic response to isoproterenol in primary rat adipocytes. *J. Mol. Endocrinol.* **42**, 57–66 (2009).
28. Fain, J. N. & Malbon, C. C. Regulation of adenylate cyclase by adenosine. *Mol. Cell. Biochem.* **25**, 143–169 (1979).
29. Miller, R. A. *et al.* Adiponectin suppresses gluconeogenic gene expression in mouse hepatocytes independent of LKB1-AMPK signaling. *J. Clin. Invest.* **121**, 2518–2528 (2011).
30. Post, S. R., Ostrom, R. S. & Insel, P. A. Biochemical methods for detection and measurement of cyclic AMP and adenylyl cyclase activity. *Methods Mol. Biol.* **126**, 363–374 (2000).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by National Institutes of Health (NIH) grants R01 DK56886 and P01 DK49210 (M.J.B.) and F32 DK079572 (R.A.M.), the Association pour l'Etude des Diabètes et des Maladies Métaboliques (ALFEDIAM) (to M.F.), the Programme National de Recherche sur le Diabète (PNRD) (to M.F. and B.V.) and the Institut Benjamin Delessert (to M.F.). Microscopy was performed in the University of Pennsylvania Cell and Developmental Biology Microscopy Core Facility. The Transgenic/Knockout, Mouse Phenotyping, Viral Vector and Biomarker Cores of the University of Pennsylvania Diabetes and Endocrinology Research Center (NIH grant P30 DK19525) were instrumental in this work.

**Author Contributions** R.A.M. and Q.C. performed experiments; R.A.M. and M.J.B. designed experiments and wrote the manuscript. J.X. generated the phospho-S33 PFKFB1 antibody. M.F. and B.V. generated the AMPK  $\alpha 1$  and  $\alpha 2$  floxed alleles. J.X., M.F. and B.V. critically read the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.B. ([birnbaum@mail.med.upenn.edu](mailto:birnbaum@mail.med.upenn.edu)).

## METHODS

**Reagents and materials.** Reconstituted GlucaGen (Novo Nordisk) was used for all glucagon studies. Phosphorylase kinase B antibody was purchased from Acris Antibodies. Total PFKFB1 antibody was purchased from Santa Cruz Biotechnology. The rabbit anti-PFKFB1 phospho-S33 antibody was produced by Cell Signaling Technology. All other antibodies were purchased from Cell Signaling Technology. The SP-8br-cAMPS-AM was purchased from AXORA, LLC. All additional reagents were purchased from Sigma Aldrich. Adeno-associated virus expressing the PKA dominant-negative Rab mutant of the PKA RI $\alpha$  subunit from a liver-specific TBG promoter was produced and purified in the University of Pennsylvania Viral Vector Core.

**Mice.** Mice were housed in a facility on a 12-h light–dark cycle with free access to food and water. All procedures were reviewed and approved by the Institutional Animal Care and Use Committee at the University of Pennsylvania. Blood glucose values were measured with OneTouch Ultra glucose analyser. Mice were killed either by rapid cervical dislocation or, for studies examining nucleotide levels, by anaesthetization with 2,2,2-tribromoethanol. Liver samples were rapidly removed and frozen in pre-cooled metal tongs and frozen in liquid nitrogen. The time from liver ischaemia to frozen tissue was approximately 30 s and 5 s for cervical dislocation or anaesthetic-based methods, respectively.

**Primary hepatocytes.** Primary hepatocytes were isolated from mice with a modified two-step perfusion method using Liver Perfusion Media and Liver Digest Buffer (Invitrogen). Cells were plated in collagen-I-coated 6- or 12-well plates (at 2 or 1 million cells per well, respectively) in M199 media plus 10% FBS plus penicillin/streptomycin plus 1 nM insulin/100 nM T3/100 nM dexamethasone. After 3 h of attachment, the media was replaced with fresh media and the cells were incubated in culture overnight. Experiments were performed 18–24 h after isolation of the cells. For 24-h treatments of hepatocytes, compounds were added after the 3-h attachment period.

**Quantification of adenine nucleotides.** Adenine nucleotides were extracted from primary hepatocytes as described previously<sup>29</sup>. For quantification of liver nucleotides, samples were harvested from anaesthetized mice and immediately frozen. Approximately 50-mg pieces of liver were weighed and nucleotides were extracted with 0.5 M perchloric acid. The insoluble materials were pelleted by centrifugation and the soluble fraction was neutralized with 0.25 volumes 2 M KOH, 1 M PO<sub>4</sub> pH 7.8. Neutralized samples were stored at –80 °C. Perchloric-acid-soluble materials were separated by isocratic elution in an ion-pair reverse-phase (RP)-HPLC system (the buffer used was 200 mM KH<sub>2</sub>PO<sub>4</sub> pH 6.25, 5 mM tetrabutylammoniumphosphate (TBAP), 3% acetonitrile). AMP, ADP and ATP peak areas were calculated and converted to molar amounts through comparison to standard curves generated from quantification of known quantities of pure nucleotides separated under identical conditions. Values of nmol mg<sup>–1</sup> protein were converted to concentrations using the cell volume of primary hepatocytes<sup>22</sup>.

**cAMP quantification.** cAMP was assayed using a cAMP ELISA kit (GE Healthcare) according to the manufacturer's alternative lysis protocol. For primary hepatocytes, cells in 12-well plates were lysed with 250  $\mu$ l of lysis buffer 1B and cAMP quantified. For liver samples, cAMP was measured by ELISA, and normalized to total liver weight.

**In vitro PKA assays.** Total soluble lysates from either primary hepatocytes or liver were obtained using lysis buffer (1% Triton X-100, 150 mM NaCl, 20 mM PO<sub>4</sub> pH 7.4). PKA activity was assayed using ATP- $\gamma$ -<sup>32</sup>P using a PKA Kinase Assay Kit (Millipore) as Kemptide phosphorylation that was sensitive to PKI inhibition.

**Microscopy.** Primary hepatocytes were plated on coverslip bottom tissue culture plates coated with collagen I and infected with adenovirus expressing the PKA FRET reporter AKAR3 (ref. 16). Eighteen hours after infection the media was changed to imaging media (M199 media lacking sodium bicarbonate and phenol red supplemented with 20 mM HEPES pH 7.4) and cells were studied using a Zeiss confocal microscope. FRET (442 nm excitation/560 nm emission), CFP (442 nm excitation/480 nm emission) and YFP (523 nm excitation/560 nm emission) channel images of hepatocytes were acquired at 20 s intervals. 5 nM glucagon was added as 500  $\mu$ l of 50 nM glucagon in pre-warmed imaging buffer to 2 ml of media. The FRET ratio ((FRET background)/(CFP background)) was taken from a random segment of the cytoplasm and nucleus, with one region of interest per cell, and over 40 regions of interest contributing to each condition. Ratios were normalized to the first 4 min of acquisition time.

**Glucose output assays.** Primary hepatocytes were isolated from 12-h fasted mice and plated in M199 media with 10% FBS for 4 h. After attachment the media was replaced with serum-free M199 media with 100 nM dexamethasone and 100 nM T3, as well as the indicated drugs. After 14 h the cells were washed with glucose output media (118 mM NaCl, 4.7 mM KCl, 1.2 mM MgSO<sub>4</sub>, 1.2 mM KH<sub>2</sub>PO<sub>4</sub>, 1.2 mM CaCl<sub>2</sub>, 20 mM NaCO<sub>3</sub>, 25 mM HEPES pH 7.4 and 0.025% BSA), and glucose output was measured in fresh glucose output media supplemented with gluconeogenic substrates (20 mM lactate, 2 mM pyruvate, 10 mM glutamine) for 5 h. Glucose output was normalized to total protein and expressed as the per cent of untreated basal glucose output.

**Adenylyl cyclase assays.** Adenylyl cyclase activity was measured as previously described<sup>30</sup>. Membrane preparations were obtained from primary hepatocytes in 15-cm plates. Cells were washed with PBS and lysed with hypotonic buffer (10 mM Tris pH 7.4, 5 mM EDTA) and 10 strokes of a dounce homogenizer. Unlysed cells and nuclei were pelleted and removed by centrifugation at 3,000g. The supernatant was centrifuged at 12,000g to pellet membranes, which were resuspended in lysis buffer and frozen at –80 °C. Adenylyl cyclase assays were performed under various ATP concentrations with tracer quantities of adenosine-5'-triphosphate [ $\alpha$ -<sup>32</sup>P] (American Radiolabelled Chemicals) with the following final assay conditions: 15 mM HEPES pH 7.4, 200 mM NaCl, 1 mM EGTA, 10 mM MgCl<sub>2</sub>, 1 mM IBMX, 10 mM phosphocreatine, 60 U ml<sup>–1</sup> creatine kinase (Sigma Aldrich). Assays were performed at 30 °C for 10 min and stopped with the addition of an equal volume of stop solution (2% SDS, 1 mM ATP, 1 mM cAMP, ~2,000 c.p.m. [<sup>3</sup>H]cAMP (refs 2, 8) (American Radiolabelled Chemicals)). Eight-hundred millilitres of water was added and this solution was purified by chromatography using Dowex 50WX8-200 (Sigma Aldrich) and Neutral Alumina (Sigma Aldrich) columns. The cAMP produced was measured as [<sup>32</sup>P]cAMP, recovered, and was corrected for recovery by calculating recovery of [<sup>3</sup>H]cAMP (refs 2, 8). The specific activity of ATP (c.p.m./total ATP) was determined before and after incubation with purified hepatocyte membrane fractions and adenylyl cyclase assay buffer in the presence and absence of AMP. No significant reduction in specific activity or total ATP was observed (Supplementary Fig. 8b, c).

**Statistical analysis.** All results are expressed as the mean  $\pm$  s.e.m. All two-group comparisons were deemed statistically significant by unpaired two-tailed Student's *t*-test if *P* < 0.05. All *in vitro* studies are either compilations of three independent experiments or representative of three independent experiments. The number of mice used in each *in vivo* condition is indicated in the figure legends.

# A T-bet gradient controls the fate and function of CCR6<sup>−</sup>RORγt<sup>+</sup> innate lymphoid cells

Christoph S. N. Klose<sup>1</sup>, Elina A. Kiss<sup>1</sup>, Vera Schwierzeck<sup>1,2</sup>, Karolina Ebert<sup>1,3</sup>, Thomas Hoyler<sup>1</sup>, Yannick d'Hargues<sup>1,3</sup>, Nathalie Göppert<sup>1</sup>, Andrew L. Croxford<sup>4,5</sup>, Ari Waisman<sup>4</sup>, Yakup Tanriver<sup>1,2,3\*</sup> & Andreas Diefenbach<sup>1,2,6\*</sup>

At mucosal surfaces, the immune system should not initiate inflammatory immune responses to the plethora of antigens constantly present in the environment, but should remain poised to unleash a potent assault on intestinal pathogens. The transcriptional programs and regulatory factors required for immune cells to switch from homeostatic (often tissue-protective) function<sup>1</sup> to potent antimicrobial immunity are poorly defined. Mucosal retinoic-acid-receptor-related orphan receptor-γt-positive (RORγt<sup>+</sup>) innate lymphoid cells (ILCs) are emerging as an important innate lymphocyte population required for immunity to intestinal infections<sup>2</sup>. Various subsets of RORγt<sup>+</sup> ILCs have been described<sup>3–8</sup> but the transcriptional programs controlling their specification and fate remain largely unknown. Here we provide evidence that the transcription factor T-bet determines the fate of a distinct lineage of CCR6<sup>−</sup>RORγt<sup>+</sup> ILCs. Postnatally emerging CCR6<sup>−</sup>RORγt<sup>+</sup> ILCs upregulated T-bet and this was controlled by cues from the commensal microbiota and interleukin-23 (IL-23). In contrast, CCR6<sup>+</sup>RORγt<sup>+</sup> ILCs, which arise earlier during ontogeny, did not express T-bet. T-bet instructed the expression of T-bet target genes such as interferon-γ (IFN-γ) and of the natural cytotoxicity receptor Nkp46. Mice genetically lacking T-bet showed normal development of CCR6<sup>−</sup>RORγt<sup>+</sup> ILCs, but they could not differentiate into Nkp46-expressing RORγt<sup>+</sup> ILCs (that is, IL-22-producing natural killer (NK-22) cells)<sup>3,9</sup> and failed to produce IFN-γ. The production of IFN-γ by T-bet-expressing CCR6<sup>−</sup>RORγt<sup>+</sup> ILCs was essential for the release of mucus-forming glycoproteins required to protect the epithelial barrier during *Salmonella enterica* infection<sup>10,11</sup>. *Salmonella* infection also causes severe enterocolitis that is at least partly driven by IFN-γ<sup>12</sup>. Mice deficient for T-bet or depleted of ILCs developed only mild enterocolitis. Thus, graded expression of T-bet in CCR6<sup>−</sup>RORγt<sup>+</sup> ILCs facilitates the differentiation of IFN-γ-producing CCR6<sup>−</sup>RORγt<sup>+</sup> ILCs required to protect the epithelial barrier against *Salmonella* infections. Co-expression of T-bet and RORγt, which is also found in subsets of IL-17-producing T-helper (T<sub>H</sub>17) cells<sup>13</sup>, may be an evolutionarily conserved transcriptional program that originally developed as part of the innate defence against infections but that also confers an increased risk of immune-mediated pathology.

Distinct subsets of RORγt<sup>+</sup> ILCs have been identified based on the expression of surface markers (for example, CD4, Nkp46 (also known as Ncr1) and stem cell growth factor receptor Kit)<sup>3–7,14</sup>, on their appearance during ontogeny<sup>6,8</sup>, and on their dependence on the transcription factor aryl hydrocarbon receptor (AhR)<sup>8,15</sup>. RORγt<sup>+</sup> ILCs continuously produce IL-22, a cytokine that fortifies the epithelial barrier<sup>3,16</sup>. In the context of intestinal inflammation, subsets of RORγt<sup>+</sup> ILCs were found to produce IFN-γ<sup>7,17</sup>. Whether these diverse functional profiles of RORγt<sup>+</sup> ILCs relate to distinct cellular subsets of RORγt<sup>+</sup> ILCs or are driven by discrete transcriptional programs is

unknown. The T-box transcription factor T-bet is a central regulator of type 1 immunity controlling the expression of IFN-γ. IFN-γ is required for powerful immunity to infections but, when inappropriately produced, can drive autoimmunity and chronic inflammation<sup>18,19</sup>. During intestinal inflammation, T-bet messenger RNA was detected in Sca-1<sup>+</sup>Thy-1<sup>high</sup> ILCs (ref. 17), some of which are RORγt<sup>+</sup> ILCs<sup>7</sup>. Unexpectedly, during steady state, one-third of all Nkp46<sup>−</sup>RORγt<sup>+</sup> ILCs of the small intestine expressed T-bet and Nkp46<sup>+</sup>RORγt<sup>+</sup> ILCs that differentiate from an Nkp46<sup>−</sup>RORγt<sup>+</sup> ILC progenitor<sup>6,7</sup> were uniformly T-bet-positive (Fig. 1a, b). Although RORγt<sup>+</sup> ILCs expressed T-bet levels comparable to natural killer cells, they lacked expression of eomesodermin, a related T-box transcription factor expressed by natural killer cells<sup>20</sup> (Supplementary Fig. 2). Thus, two populations of RORγt<sup>+</sup> ILCs can be discriminated based on T-bet expression.

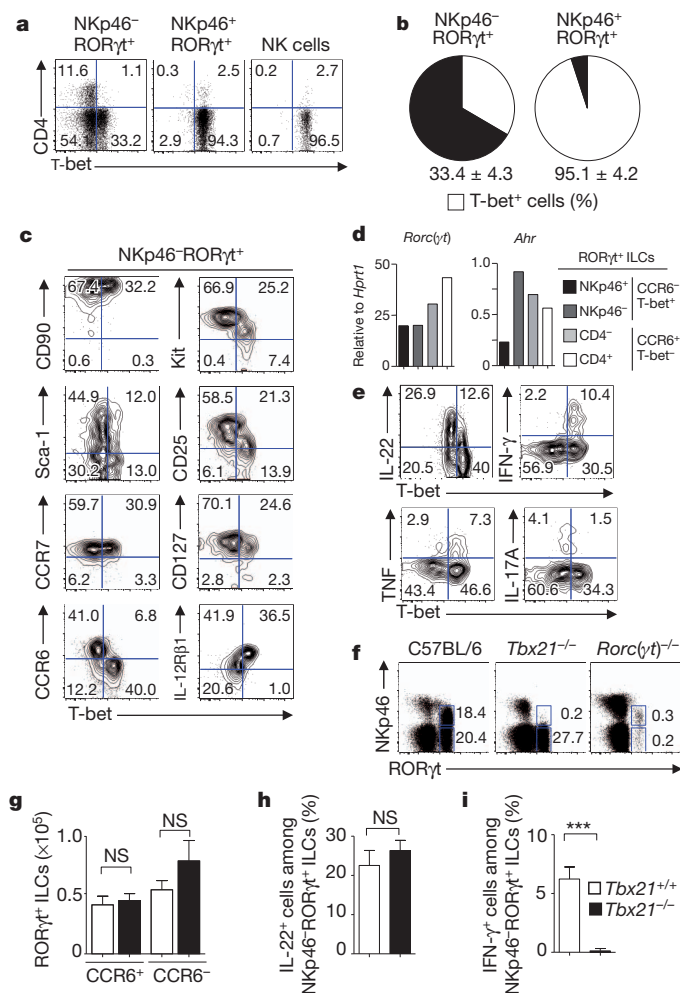
Both T-bet<sup>−</sup> and T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs expressed comparable levels of *Rorc*(γt) (the gene encoding RORγt), *Ahr*, and of Thy-1 (also known as CD90), Sca-1 and CCR7 (CC-chemokine receptor type 7) commonly expressed by ILCs (Fig. 1c, d and Supplementary Fig. 3a). Compared to T-bet<sup>−</sup>RORγt<sup>+</sup> ILCs, both Nkp46<sup>−</sup> and Nkp46<sup>+</sup>T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs had lower expression of Kit, IL-2Rα (also known as CD25) and IL-7Rα (also known as CD127) and higher expression of IL-12Rβ1, a subunit of the IL-12 and IL-23 receptors. Remarkably, Nkp46<sup>−</sup> and Nkp46<sup>+</sup>T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs did not express CCR6, whereas the majority of T-bet<sup>−</sup>RORγt<sup>+</sup> ILCs were CCR6<sup>+</sup> (Fig. 1c and Supplementary Fig. 3a, b). Transcriptional profiling confirmed that *Tbx21*, the gene encoding T-bet, is selectively expressed by CCR6<sup>−</sup>RORγt<sup>+</sup> ILCs that also express various T-bet target genes<sup>19</sup>, which were absent from CCR6<sup>+</sup>T-bet<sup>−</sup>RORγt<sup>+</sup> ILCs (Supplementary Fig. 3c). Interestingly, T-bet expression in CCR6<sup>−</sup>RORγt<sup>+</sup> ILCs increased from the Nkp46<sup>−</sup> to the Nkp46<sup>+</sup> subset (Supplementary Fig. 3d). CCR6<sup>+</sup>T-bet<sup>−</sup> and CCR6<sup>−</sup>T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs expressed comparable levels of various TNF superfamily genes required for lymphoid-tissue-inducing function (Supplementary Fig. 4a). Both subsets produced IL-22, but at different levels (Fig. 1e and Supplementary Fig. 4b, c). CCR6<sup>−</sup>T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs produced IFN-γ and TNF, whereas CCR6<sup>+</sup>T-bet<sup>−</sup>RORγt<sup>+</sup> ILCs expressed IL-17A and IL-17F (Fig. 1e and Supplementary Fig. 4b).

Development of CCR6<sup>+</sup> and Nkp46<sup>−</sup>CCR6<sup>−</sup>RORγt<sup>+</sup> ILCs was not perturbed in mice genetically lacking T-bet<sup>21</sup> (Fig. 1f, g). Notably, the Nkp46<sup>+</sup> population of CCR6<sup>−</sup>RORγt<sup>+</sup> ILCs (that is, NK-22 cells)<sup>9</sup> was absent in *Tbx21*<sup>−/−</sup> mice (Fig. 1f and Supplementary Fig. 5a), demonstrating that T-bet is required for the differentiation of Nkp46<sup>+</sup>RORγt<sup>+</sup> ILCs from their Nkp46<sup>−</sup> progenitors<sup>6,7</sup>. Experiments with mixed bone-marrow chimerae revealed that T-bet acts in an RORγt<sup>+</sup> ILC-intrinsic fashion (Supplementary Fig. 5b, c). In addition to lacking Nkp46 expression, T-bet-deficient RORγt<sup>+</sup> ILCs

<sup>1</sup>Institute of Medical Microbiology and Hygiene (IMMH), University of Freiburg, Hermann-Herder-Strasse 11, D-79104 Freiburg, Germany. <sup>2</sup>Centre of Chronic Immunodeficiency (CCI), University Medical Center Freiburg and University of Freiburg, Breisacher Strasse 117, D-79106 Freiburg, Germany. <sup>3</sup>Renal Division, University of Freiburg Medical Centre, Hugstetter Strasse 55, D-79106 Freiburg, Germany. <sup>4</sup>Institute of Molecular Medicine, University of Mainz, Obere Zahlbacher Strasse 67, D-55131 Mainz, Germany. <sup>5</sup>Institute of Experimental Immunology, Neuroimmunology, University of Zürich, CH-8006 Zürich, Switzerland. <sup>6</sup>Centre for Biological Signalling Studies (BIOS), University of Freiburg, Schänzlestrasse 18, D-79104 Freiburg, Germany.

\*These authors contributed equally to this work.





**Figure 1** | CCR6<sup>+</sup> RORγt<sup>+</sup> ILCs express T-bet. **a**, Flow cytometry analysis of T-bet and CD4 expression by the indicated innate lymphocyte subsets of the small intestine (gating strategy is given in Supplementary Fig. 1). **b**, Percentage (mean ± s.d.;  $n = 12$ ) of T-bet<sup>+</sup> cells within NKp46<sup>-</sup> RORγt<sup>+</sup> and NKp46<sup>+</sup> RORγt<sup>+</sup> ILCs. **c**, Co-expression of T-bet with the indicated cell-surface markers by NKp46<sup>-</sup> RORγt<sup>+</sup> ILCs. **d**, Expression of the indicated genes was analysed by qRT-PCR using complementary DNA from the indicated RORγt<sup>+</sup> ILC subsets. **e**, T-bet and cytokine expression by intestinal NKp46<sup>-</sup> RORγt<sup>+</sup> ILCs after phorbol myristate acetate (PMA) and ionomycin stimulation. **f**, NKp46 and RORγt expression by Lin<sup>-</sup> CD45<sup>+</sup> lamina propria lymphocytes of the indicated mouse strains. **g**, Absolute numbers (mean ± s.d.;  $n = 6$ ) of NKp46<sup>-</sup> RORγt<sup>+</sup> ILC subsets. **h**, **i**, Percentage (mean ± s.d.;  $n = 3$ ) of PMA- and ionomycin-stimulated NKp46<sup>+</sup> RORγt<sup>+</sup> ILCs expressing the indicated cytokines. Numbers represent percentage of cells per gate or quadrant. Data are representative of three independent experiments. \*\*\* $P < 0.001$ ; NK, natural killer; NS, not significant.

did not express IFN-γ and other T-bet target genes, whereas IL-22 and IL-17 were unaffected (Fig. 1h, i and Supplementary Fig. 5d, e).

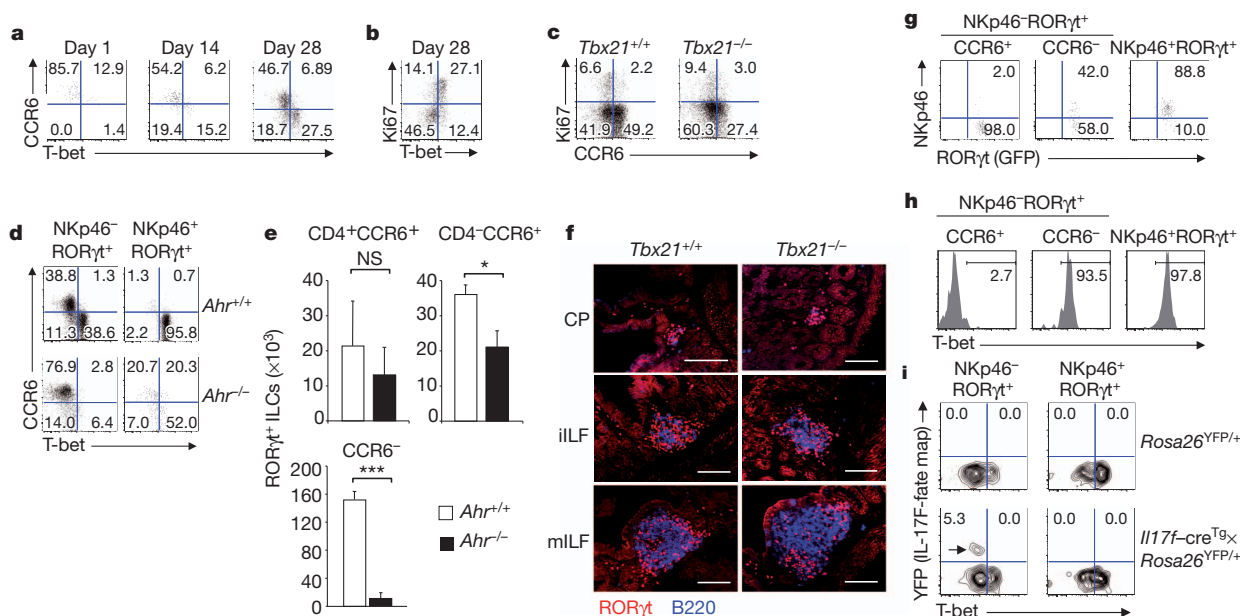
Expression of T-bet exclusively in CCR6<sup>+</sup> RORγt<sup>+</sup> ILCs, in conjunction with substantially increased numbers of CCR6<sup>+</sup> RORγt<sup>+</sup> ILCs in *Tbx21*<sup>-/-</sup> mice and deficiency of NKp46<sup>+</sup> CCR6<sup>+</sup> RORγt<sup>+</sup> ILCs raised important questions about the lineage relationship between T-bet<sup>+</sup> RORγt<sup>+</sup> ILCs and other RORγt<sup>+</sup> ILC populations. Both T-bet<sup>+</sup> and T-bet<sup>-</sup> subsets of RORγt<sup>+</sup> ILCs expressed RORγt and were absent in mice genetically lacking RORγt expression (Fig. 1d, f), confirming that they are true RORγt<sup>+</sup> ILCs. Preferential expression of T-bet in CD4<sup>-</sup> RORγt<sup>+</sup> ILCs was interesting in the light of our previous data, assigning an important role to the transcription factor AhR for the postnatal expansion of CD4<sup>-</sup> RORγt<sup>+</sup> ILCs<sup>8</sup>. The AhR-dependent subset of RORγt<sup>+</sup> ILCs expands during weeks 2 and 3 after birth in a

process that requires proliferation<sup>8</sup>. CCR6<sup>+</sup> T-bet<sup>-</sup> RORγt<sup>+</sup> ILCs were the dominant RORγt<sup>+</sup> ILC population during the first 2 weeks after birth (Fig. 2a). Thereafter, CCR6<sup>-</sup> RORγt<sup>+</sup> ILCs appeared and the fraction of T-bet-expressing CCR6<sup>+</sup> RORγt<sup>+</sup> ILCs increased with age (Fig. 2a–c). Notably, the majority of CCR6<sup>-</sup> T-bet<sup>+</sup> RORγt<sup>+</sup> ILCs of 4-week-old mice expressed Ki67, a marker strictly associated with cellular proliferation (Fig. 2b and Supplementary Fig. 6a, b). Proliferation was, however, not dependent on T-bet (Fig. 2c). We investigated the representation of CCR6<sup>+</sup> and CCR6<sup>-</sup> RORγt<sup>+</sup> ILCs in AhR-deficient mice. Although *Ahr*<sup>-/-</sup> mice had mainly normal numbers of CD4<sup>+</sup> CCR6<sup>+</sup> T-bet<sup>-</sup> RORγt<sup>+</sup> ILCs, and CD4<sup>-</sup> CCR6<sup>+</sup> T-bet<sup>-</sup> RORγt<sup>+</sup> ILCs were reduced less than twofold, both the T-bet<sup>-</sup> and T-bet<sup>+</sup> subsets of CCR6<sup>-</sup> RORγt<sup>+</sup> ILCs were dramatically diminished in *Ahr*<sup>-/-</sup> mice (Fig. 2d, e). Thus, postnatally emerging CCR6<sup>-</sup> RORγt<sup>+</sup> ILCs are the AhR-dependent subset of RORγt<sup>+</sup> ILCs, the presence of which is required for the formation of cryptopatches and isolated lymphoid follicles<sup>8</sup>. In contrast to *Ahr*<sup>-/-</sup> mice (Fig. 2e), *Tbx21*<sup>-/-</sup> mice had a normal number of CCR6<sup>-</sup> RORγt<sup>+</sup> ILCs (Fig. 1g) expressing high levels of genes implicated in lymphoid-tissue-inducing function (Supplementary Fig. 4a) and hence showed normal development of cryptopatches and isolated lymphoid follicles (Fig. 2f).

Our findings indicate that CCR6<sup>+</sup> and CCR6<sup>-</sup> RORγt<sup>+</sup> ILCs may constitute separate lineages and that CCR6<sup>-</sup> T-bet<sup>-</sup> RORγt<sup>+</sup> ILCs differentiate into CCR6<sup>-</sup> T-bet<sup>+</sup> RORγt<sup>+</sup> ILCs that then acquire NKp46. To test this directly, we transferred genetically labelled subsets of RORγt<sup>+</sup> ILCs into alymphoid mice. Donor-derived CCR6<sup>+</sup> RORγt<sup>+</sup> ILCs remained RORγt<sup>+</sup> and did not acquire T-bet or NKp46 expression (Fig. 2g, h and Supplementary Fig. 7a), a finding that was confirmed with *in vitro* differentiation assays (Supplementary Fig. 7b). Stability of CCR6<sup>+</sup> RORγt<sup>+</sup> ILCs was supported further by the analysis of an IL-17F fate map (*Il17f-cre*<sup>TB</sup> × *Rosa26*<sup>YFP/+</sup>). Yellow fluorescent protein (YFP)-positive (YFP<sup>+</sup>; IL-17F-fate-map-positive) cells could only be found in the CCR6<sup>+</sup> T-bet<sup>-</sup> fraction of RORγt<sup>+</sup> ILCs, demonstrating that IL-17F-expressing CCR6<sup>+</sup> RORγt<sup>+</sup> ILCs did not differentiate into T-bet<sup>+</sup> cells (Fig. 2i). Chronic inflammation may also induce T-bet expression in IL-17-producing ILCs<sup>17</sup>. Although CCR6<sup>+</sup> RORγt<sup>+</sup> ILCs were largely stable, NKp46<sup>-</sup> CCR6<sup>-</sup> RORγt<sup>+</sup> ILCs first upregulated T-bet expression and then acquired NKp46 both *in vivo* (Fig. 2g, h) and *in vitro* (Supplementary Fig. 7b). Consistent with previous data<sup>7,22</sup>, NKp46<sup>+</sup> RORγt<sup>+</sup> ILCs downregulated RORγt expression, enabling functional plasticity (Supplementary Fig. 7b).

Collectively, our data provide evidence for two lineages of IL-22-producing RORγt<sup>+</sup> ILCs that can be discriminated on the basis of CCR6 expression. CCR6<sup>+</sup> RORγt<sup>+</sup> ILCs seed the lamina propria of the small intestine during fetal development, do not require AhR for maintenance and do not express T-bet or upregulate NKp46. Similar to fetal lymphoid tissue inducer cells<sup>14</sup>, only CCR6<sup>+</sup> RORγt<sup>+</sup> ILCs expressed IL-17A and IL-17F. CCR6<sup>-</sup> RORγt<sup>+</sup> ILCs expand after birth under the control of AhR signals and can upregulate T-bet, which is required for IFN-γ expression and the differentiation of NKp46<sup>+</sup> RORγt<sup>+</sup> ILCs. Previous work has proposed that lineages of RORγt<sup>+</sup> ILCs can be discriminated based on the levels of Kit and IL-7Rα expression<sup>6</sup>. Interestingly, T-bet repressed Kit and IL-7Rα expression in CCR6<sup>-</sup> RORγt<sup>+</sup> ILCs. CCR6<sup>-</sup> RORγt<sup>+</sup> ILCs from *Tbx21*<sup>-/-</sup> mice (Supplementary Fig. 8a) and the small subset of CCR6<sup>-</sup> T-bet<sup>-</sup> RORγt<sup>+</sup> ILCs (Supplementary Fig. 8b) in wild-type mice had Kit and IL-7Rα levels comparable to the CCR6<sup>+</sup> subset. Thus, 'Kit<sup>high</sup>' RORγt<sup>+</sup> ILCs<sup>6</sup> are a mixed population of CCR6<sup>+</sup> RORγt<sup>+</sup> ILCs that arise fetally and CCR6<sup>-</sup> T-bet<sup>-</sup> RORγt<sup>+</sup> ILCs that expand postnatally.

If CCR6<sup>-</sup> T-bet<sup>-</sup> RORγt<sup>+</sup> ILCs are the progenitors of CCR6<sup>-</sup> T-bet<sup>+</sup> RORγt<sup>+</sup> ILCs, the question arises as to what signals upregulate T-bet expression. Germ-free mice have reduced numbers of NKp46<sup>+</sup> RORγt<sup>+</sup> ILCs<sup>3,4,7</sup> and we reasoned that the commensal microbiota may instruct T-bet expression. Indeed, germ-free mice had significantly lower numbers of T-bet-expressing NKp46<sup>+</sup> RORγt<sup>+</sup> ILCs



**Figure 2** | CCR6<sup>-</sup>T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs and CCR6<sup>+</sup>T-bet<sup>-</sup>RORγt<sup>+</sup> ILCs are separate ILC lineages. **a–c**, Expression of T-bet and CCR6 (**a**), T-bet and Ki67 (**b**), or CCR6 and Ki67 (**c**) by intestinal NKp46<sup>-</sup>RORγt<sup>+</sup> ILCs at the indicated time points after birth. **d**, Co-expression of CCR6 and T-bet by intestinal NKp46<sup>-</sup> and NKp46<sup>+</sup>RORγt<sup>+</sup> ILCs from 8-week-old mice. **e**, Absolute numbers (mean ± s.d.; *n* = 3) of the indicated NKp46<sup>-</sup>RORγt<sup>+</sup> ILC subsets. **f**, Intestinal lymphoid follicles in the small intestine of 8-week-old mice. Scale bar, 100 μm. Original magnification, ×20. CP, cryptopatch; iILF,

immature isolated lymphoid follicle; mLIF, mature ILF. **g, h**, Highly purified RORγt<sup>+</sup> ILC subsets isolated from *Rorc*(γt)<sup>GFP/+</sup> mice (H-2<sup>b</sup>) were adoptively transferred into lymphoid *Rag2*<sup>-/-</sup>*Il2rg*<sup>-/-</sup> mice (H-2<sup>d</sup>). Three weeks later, donor-derived cells were analysed for the expression of NKp46 and RORγt (green fluorescent protein; GFP) (**g**) or T-bet (**h**). **i**, T-bet- and IL-17F-fate-label expression by intestinal RORγt<sup>+</sup> ILC subsets. Numbers represent percentage of cells in quadrants. Arrow highlights the IL-17F-fate label-positive cells. \**P* < 0.05; \*\*\**P* < 0.001; NS, not significant.

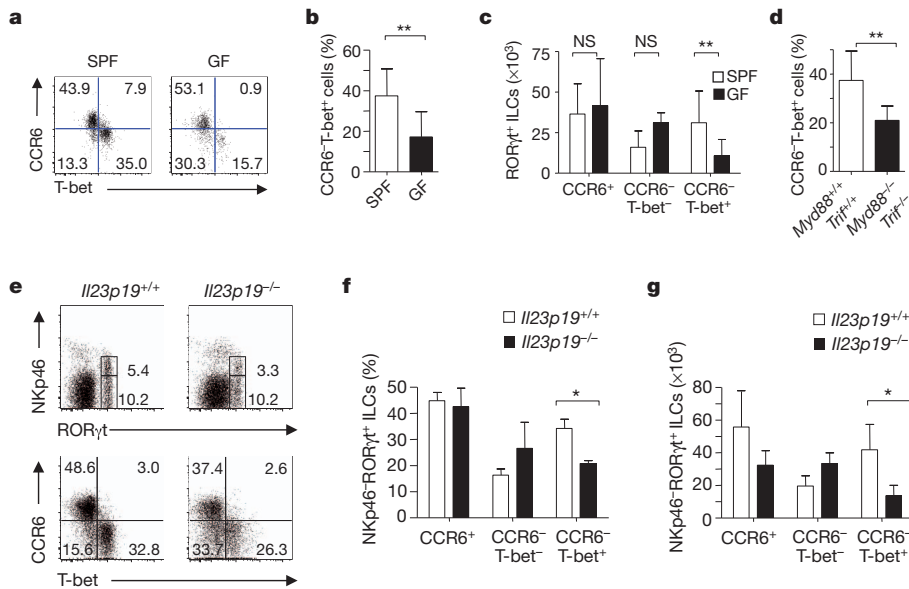
than conventionally raised, specific-pathogen-free (SPF) mice and a corresponding increase in NKp46<sup>+</sup>CCR6<sup>-</sup>T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs (Fig. 3a–c). Microbial cues required for T-bet expression by CCR6<sup>-</sup>RORγt<sup>+</sup> ILCs may be generated by Toll-like receptors (TLRs) because mice lacking TLR signalling (that is, *Myd88*<sup>-/-</sup>*Trif*<sup>-/-</sup> mice) showed a similar reduction in CCR6<sup>-</sup>T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs (Fig. 3d). This does not reflect defective IL-1 or TLR signalling in RORγt<sup>+</sup> ILCs<sup>23</sup>, because mice with a tissue-specific deletion of the *Myd88* gene in all RORγt<sup>+</sup> ILCs have normal numbers of CCR6<sup>-</sup>T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs (Supplementary Fig. 9a–c). The IL-12 family of cytokines has an important role in the function and differentiation of RORγt<sup>+</sup> ILCs<sup>7</sup>. Mice lacking principal components of IL-12 signalling (*Il12rb2*<sup>-/-</sup> or *Stat4*<sup>-/-</sup> mice) had normal numbers of CCR6<sup>-</sup>T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs (Supplementary Fig. 10a–d). In contrast, mice lacking IL-23 (*Il23p19*<sup>-/-</sup> mice) had reduced numbers of T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs (Fig. 3e–g), whereas the number of NKp46<sup>+</sup>RORγt<sup>+</sup> cells was normal (Supplementary Fig. 10e). Interestingly, IL-23 levels in various parts of the small intestine (Supplementary Fig. 11a) correlated with the fraction of T-bet<sup>+</sup> cells among CCR6<sup>-</sup>RORγt<sup>+</sup> ILCs (Supplementary Fig. 11b–d). Thus, microbiota and IL-23, in part, control the expression of T-bet in CCR6<sup>-</sup>RORγt<sup>+</sup> ILCs.

Ahr-dependent and IL-22-producing RORγt<sup>+</sup> ILCs are essential for the defence against intestinal attaching-and-effacing infections such as those caused by *Citrobacter rodentium* in mice<sup>8,24,25</sup>. As *Tbx21*<sup>-/-</sup> mice had normal numbers of IL-22-producing RORγt<sup>+</sup> ILCs (Supplementary Fig. 12a) and normal epithelial expression of the *Reg3g* gene (encoding an antimicrobial lectin)<sup>26</sup> (Supplementary Fig. 12b), they were resistant to infection with *C. rodentium* (Supplementary Fig. 12c, d). These data also show that NKp46<sup>+</sup>T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs, absent in *Tbx21*<sup>-/-</sup> mice, do not have a leading role in the defence against *C. rodentium* infection. Previous work has established that mice deficient for both *Rag* (lacking all T and B cells) and *Tbx21* develop spontaneous colitis (*Tbx21*<sup>-/-</sup>*Rag*<sup>-/-</sup> ulcerative colitis, TRUC)<sup>27</sup>. In *Tbx21*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice, intestinal mononuclear phagocytes produced high levels of TNF and contained pathogenic

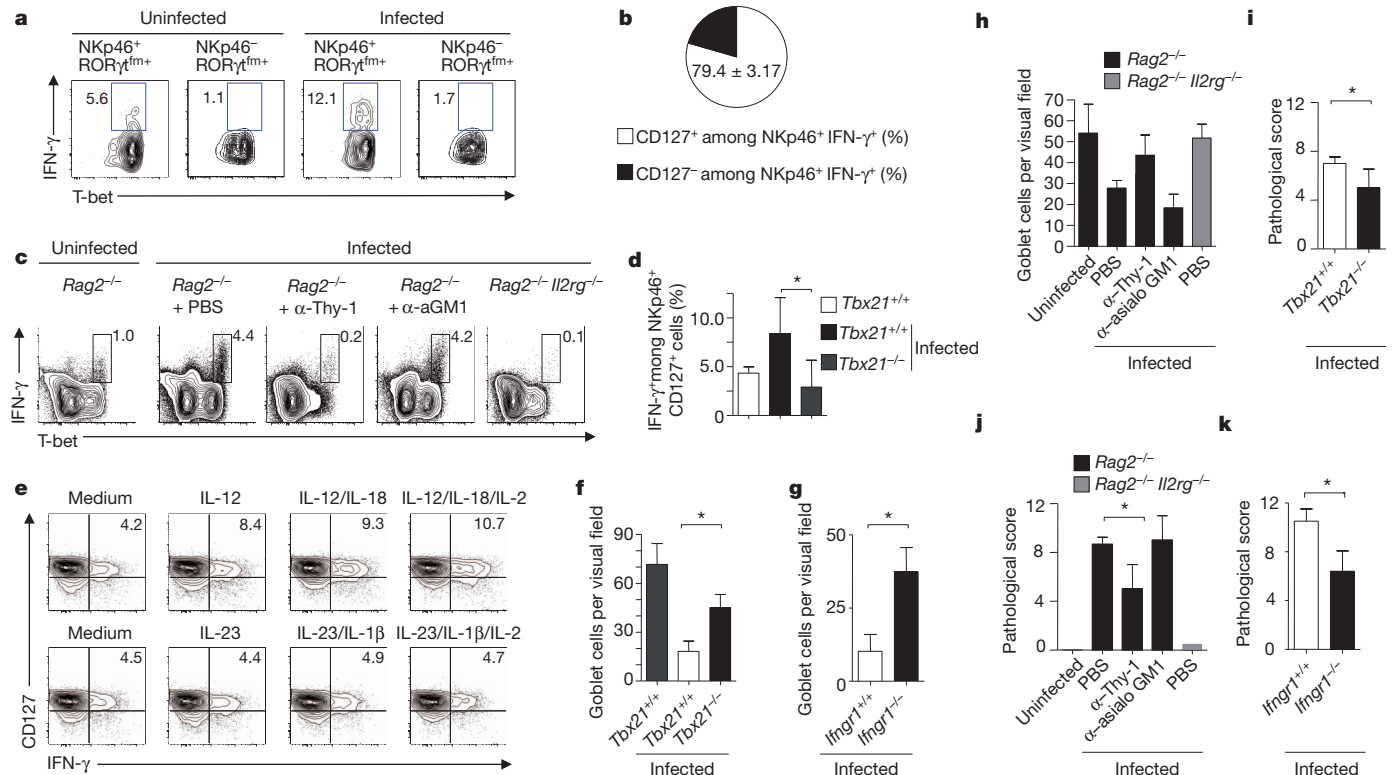
microbial communities characterized by the outgrowth of Gram-negative Enterobacteriaceae<sup>27</sup>. However, these inflammatory changes were not found in T- and B-cell-proficient *Tbx21*<sup>-/-</sup> mice, excluding them as confounding factors in our experiments (Supplementary Figs 13 and 14).

We reasoned that T-bet-controlled IFN-γ production by CCR6<sup>-</sup>RORγt<sup>+</sup> ILCs may be an important factor in other types of frequent intestinal infections such as *Salmonella enterica* (serovar Typhimurium) infections, resistance to which requires IFN-γ production<sup>10,11</sup>. We observed that NKp46<sup>+</sup>T-bet<sup>+</sup>RORγt-fate-map-positive ILCs (most of which had low RORγt expression), but not NKp46<sup>-</sup>T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs, were the main source of IFN-γ in response to *S. typhimurium* infection (Fig. 4a, b). In all parts of the intestine, the majority of IFN-γ-producers were CD127<sup>+</sup> ILCs (containing RORγt-fate-map-positive ILCs and putative non-NK ILC1)<sup>2</sup>, whereas only a small fraction were CD127<sup>-</sup> natural killer cells (Fig. 4b and Supplementary Fig. 15a). Selective depletion<sup>7</sup> of ILCs (α-Thy-1) but not of natural killer cells (α-asialo-GM1) eliminated *Salmonella*-induced IFN-γ production in *Rag2*<sup>-/-</sup> mice (Fig. 4c). As *Tbx21*<sup>-/-</sup> mice lacked NKp46<sup>+</sup>RORγt<sup>+</sup> ILCs, no infection-induced expression of IFN-γ by ILCs was observed (Fig. 4d and Supplementary Fig. 15b). Interestingly, IL-12 but not IL-23 induced IFN-γ production by NKp46<sup>+</sup>T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs (Fig. 4e), and this is consistent with the finding that IL-12-deficient mice, but not IL-23-deficient mice, lack IFN-γ production after *Salmonella* infection<sup>28</sup>.

IFN-γ produced in response to *Salmonella* has diverse roles. IFN-γ controls mucin release by goblet cells, and this enables the formation of a mucin gel that protects the epithelial barrier<sup>10</sup>. Significantly fewer epithelial cells containing mucus vacuoles were discernable in *Salmonella*-infected *Tbx21*<sup>+/+</sup> mice than in uninfected *Tbx21*<sup>+/+</sup> mice, demonstrating that their goblet cells had secreted mucus-forming glycoproteins. *Tbx21*<sup>-/-</sup> mice lacking IFN-γ-producing NKp46<sup>+</sup>RORγt<sup>+</sup> ILCs showed a significant defect in the ability to secrete mucin (Fig. 4f and Supplementary Fig. 15c). Notably, *Rag2*<sup>-/-</sup> mice that were selectively depleted of IFN-γ-producing ILCs



**Figure 3 | Microbiota and IL-23 control T-bet expression by CCR6<sup>-</sup>RORγt<sup>+</sup> ILCs.** **a**, T-bet and CCR6 expression by NKp46<sup>+</sup>RORγt<sup>+</sup> ILCs from conventionally housed (SPF) or germ-free (GF) mice. **b**, Percentage (mean + s.d.; *n* = 12) of cells among NKp46<sup>+</sup>RORγt<sup>+</sup> ILCs from SPF or GF mice. **c**, Absolute cell numbers (mean + s.d.; *n* = 12) of NKp46<sup>+</sup>RORγt<sup>+</sup> ILCs within the indicated populations. **d**, Percentage (mean + s.d.; *n* = 10) of CCR6<sup>+</sup>T-bet<sup>+</sup> cells among NKp46<sup>+</sup>RORγt<sup>+</sup> ILCs from Myd88<sup>-/-</sup>Trif<sup>-/-</sup> mice or control mice. **e**, NKp46 and RORγt expression of Lin<sup>-</sup> lamina propria lymphocytes from the small intestine of Il23p19<sup>-/-</sup> or control mice (top panels). CCR6 and T-bet expression of NKp46<sup>+</sup>RORγt<sup>+</sup> ILCs is shown in the lower panels. **f**, **g**, Percentages (**f**) and absolute cell numbers (**g**) (mean + s.d.; *n* = 3) of the indicated subsets among NKp46<sup>+</sup>RORγt<sup>+</sup> ILCs. Numbers represent percentage of cells in gates or quadrants. Data are representative of at least two independent experiments. \**P* < 0.05; \*\**P* < 0.01; NS, not significant.



**Figure 4 | IFN-γ-producing NKp46<sup>+</sup>CCR6<sup>-</sup>T-bet<sup>+</sup>RORγt<sup>+</sup> ILCs control mucus release and inflammation following *Salmonella* infection.**

**a**, **b**, Groups of *Rorc*(γt)-*cre*<sup>Tg</sup>×*Rosa26R*<sup>YFP/+</sup> (RORγt fate map, RORγt<sup>fm</sup>) mice were infected with *S. typhimurium*. One day later, IFN-γ production in electronically gated Lin<sup>-</sup> NKp46<sup>+</sup> or NKp46<sup>-</sup>RORγt<sup>fm</sup> ILCs from the lamina propria of the small intestine was determined after 4 h stimulation with IL-12 (10 ng ml<sup>-1</sup>) (**a**). Percentage (mean ± s.d.; *n* = 5) of CD127<sup>+</sup> ILCs among IFN-γ<sup>+</sup> NKp46<sup>+</sup> lymphocytes (**b**). **c**, Groups of *Rag2*<sup>-/-</sup> and *Rag2*<sup>-/-</sup> *Il2rg*<sup>-/-</sup> mice were infected with *S. typhimurium* and treated with the indicated antibody or PBS. Control mice were left uninfected. After 96 h, IFN-γ and T-bet expression was determined. **d**, Groups of *Tbx21*<sup>+/+</sup> and *Tbx21*<sup>-/-</sup> mice were infected with *S. typhimurium*. Control mice were left uninfected. After 48 h, the percentage

(mean ± s.d.; *n* = 5) of IFN-γ<sup>+</sup> cells among NKp46<sup>+</sup> CD127<sup>+</sup> ILCs was determined. **e**, IFN-γ and CD127 expression by lamina propria lymphocytes from *Rag2*<sup>-/-</sup> mice infected for 96 h with *S. typhimurium* and after 4 h stimulation with the indicated cytokines. **f**–**k**, Groups of the indicated mouse strains were infected with *S. typhimurium* and treated with the indicated antibodies. Two or four days after infection (*Rag2*<sup>-/-</sup> mice), the number of goblet cells per visual field were counted in H&E-stained sections (**f**–**h**) and pathological scores quantifying the extent of inflammation (**i**–**k**) were determined. Numbers represent percentage of cells in gates. Data are representative of at least two independent experiments. \**P* < 0.05; α-aGM1, α-aGM1.



could not secrete mucin proteins, similar to mice lacking the *Ifngr* gene (Fig. 4g, h). Depletion of natural killer cells had no bearing on mucus secretion. The production of IFN- $\gamma$  to protect the epithelial barrier against *Salmonella* can also promote enterocolitis. Therefore, we sought to determine whether IFN- $\gamma$ -producing CCR6<sup>+</sup>T-bet<sup>+</sup>ROR $\gamma$ t<sup>+</sup>-fate-map-positive ILCs were involved in the pathogenesis of enterocolitis. *Tbx21*<sup>-/-</sup> mice (Fig. 4i and Supplementary Fig. 15d–f) and ILC-depleted mice (Fig. 4j and Supplementary Fig. 16a–d) infected with *S. typhimurium* showed decreased clinical signs of enterocolitis, similar to *Ifngr*<sup>-/-</sup> mice (Fig. 4k and Supplementary Fig. 16e–g).

Co-expression of T-bet and ROR $\gamma$ t has been found under inflammatory conditions in TH17 cells, a lymphocyte subset with similar transcriptional circuitry and effector profiles to ROR $\gamma$ t<sup>+</sup> ILCs<sup>13,29,30</sup>. Such T-bet-expressing TH17 cells promote autoimmunity and chronic inflammation. We show here that this transcriptional circuitry is pre-formed in the innate immune system and may have evolved to aid the innate defence against infections. Notably, the majority of intestinal CCR6<sup>+</sup>ROR $\gamma$ t<sup>+</sup> ILCs co-expressed T-bet and ROR $\gamma$ t during steady state and the extent of T-bet expression was controlled by the environment (Supplementary Fig. 17). Our data and related data regarding human ILCs<sup>31</sup> reveal that induced and tunable co-expression of T-bet in CCR6<sup>+</sup>ROR $\gamma$ t<sup>+</sup> ILCs serves as a molecular switch to control a gene-expression program that enables a homeostatic phenotype (that is, IL-22) to switch to a proinflammatory phenotype in response to extracellular intestinal pathogens such as *Salmonella*.

## METHODS SUMMARY

A detailed list of all genetically modified mouse strains used in this study can be found in the Methods. Isolation of lamina propria leukocytes was carried out as described previously<sup>8</sup>. Single-cell suspensions of lamina propria leukocytes were from the small intestine (unless explicitly stated) and were analysed by flow cytometry or sorted using a BD FACSARIA III cell sorter. For adoptive transfer experiments, highly purified lymphocytes (>98%) were injected intravenously into *Rag2*<sup>-/-</sup>*Il2rg*<sup>-/-</sup> mice on a BALB/c background. For quantitative polymerase chain reaction with reverse transcription (qRT-PCR) analysis, data are presented as *n*-fold difference to the gene *Hprt1*. All immunofluorescence stainings were carried out on ice and after blocking of Fc receptors with an antibody specific for CD16 and CD32. Lineage-positive cells were excluded by staining with antibodies specific for CD3, CD5, CD8 $\alpha$ , CD19 and Ly6C and Ly6G. For intracellular cytokine staining, single-cell suspensions of the indicated cells were incubated in medium with brefeldin A for 4 h at 37 °C. Indicated cytokines or PMA and ionomycin were added to some wells in order to trigger cytokine production. Infections with *C. rodentium* ICC169 or *S. enterica* serovar typhimurium strain SL1344 (SB300) were carried out as described previously<sup>8,12</sup>. For histology, tissues were fixed in 4% paraformaldehyde, washed in PBS, embedded in optimal cutting temperature (OCT) compound and snap-frozen in liquid nitrogen. Histological sections of the ileum or caecum were stained with antibodies against the indicated epitopes or with haematoxylin and eosin (H&E). Data were calculated as mean  $\pm$  standard deviation (s.d.). To improve clarity, only mean + s.d. is shown in the figures.

**Full Methods** and any associated references are available in the online version of the paper.

Received 6 September; accepted 26 November 2012.

Published online 16 January 2013.

- Hooper, L. V. & Macpherson, A. J. Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nature Rev. Immunol.* **10**, 159–169 (2010).
- Spits, H. & Cupedo, T. Innate lymphoid cells: emerging insights in development, lineage relationships, and function. *Annu. Rev. Immunol.* **30**, 647–675 (2012).
- Sanos, S. L. *et al.* ROR $\gamma$ t and commensal microflora are required for the differentiation of mucosal interleukin 22-producing NKp46<sup>+</sup> cells. *Nature Immunol.* **10**, 83–91 (2009).
- Satoh-Takayama, N. *et al.* Microbial flora drives interleukin 22 production in intestinal NKp46<sup>+</sup> cells that provide innate mucosal immune defense. *Immunity* **29**, 958–970 (2008).
- Luci, C. *et al.* Influence of the transcription factor ROR $\gamma$ t on the development of NKp46<sup>+</sup> cell populations in gut and skin. *Nature Immunol.* **10**, 75–82 (2009).
- Sawa, S. *et al.* Lineage relationship analysis of ROR $\gamma$ t<sup>+</sup> innate lymphoid cells. *Science* **330**, 665–669 (2010).

- Vonarbourg, C. *et al.* Regulated expression of nuclear receptor ROR $\gamma$ t confers distinct functional fates to NK cell receptor-expressing ROR $\gamma$ t<sup>+</sup> innate lymphocytes. *Immunity* **33**, 736–751 (2010).
- Kiss, E. A. *et al.* Natural aryl hydrocarbon receptor ligands control organogenesis of intestinal lymphoid follicles. *Science* **334**, 1561–1565 (2011).
- Cella, M. *et al.* A human natural killer cell subset provides an innate source of IL-22 for mucosal immunity. *Nature* **457**, 722–725 (2009).
- Songhet, P. *et al.* Stromal IFN- $\gamma$ -signaling modulates goblet cell function during *Salmonella* Typhimurium infection. *PLoS ONE* **6**, e22459 (2011).
- Rhee, S. J., Walker, W. A. & Cherayil, B. J. Developmentally regulated intestinal expression of IFN- $\gamma$  and its target genes and the age-specific response to enteric *Salmonella* infection. *J. Immunol.* **175**, 1127–1136 (2005).
- Barthel, M. *et al.* Pretreatment of mice with streptomycin provides a *Salmonella enterica* serovar Typhimurium colitis model that allows analysis of both pathogen and host. *Infect. Immun.* **71**, 2839–2858 (2003).
- Ghoreschi, K. *et al.* Generation of pathogenic TH17 cells in the absence of TGF- $\beta$  signalling. *Nature* **467**, 967–971 (2010).
- Cupedo, T. *et al.* Human fetal lymphoid tissue-inducer cells are interleukin 17-producing precursors to ROR $\gamma$ t<sup>+</sup> CD127<sup>+</sup> natural killer-like cells. *Nature Immunol.* **10**, 66–74 (2009).
- Lee, J. S. *et al.* AHR drives the development of gut ILC22 cells and postnatal lymphoid tissues via pathways dependent on and independent of Notch. *Nature Immunol.* **13**, 144–151 (2011).
- Sanos, S. L., Vonarbourg, C., Mortha, A. & Diefenbach, A. Control of epithelial cell function by interleukin-22-producing ROR $\gamma$ t<sup>+</sup> innate lymphoid cells. *Immunology* **132**, 453–465 (2011).
- Buonocore, S. *et al.* Innate lymphoid cells drive interleukin-23-dependent innate intestinal pathology. *Nature* **464**, 1371–1375 (2010).
- Oestreich, K. J. & Weinmann, A. S. Transcriptional mechanisms that regulate T helper 1 cell differentiation. *Curr. Opin. Immunol.* **24**, 191–195 (2012).
- Lazarevic, V. & Glimcher, L. H. T-bet in disease. *Nature Immunol.* **12**, 597–606 (2011).
- Gordon, S. M. *et al.* The transcription factors T-bet and Eomes control key checkpoints of natural killer cell maturation. *Immunity* **36**, 55–67 (2012).
- Szabo, S. J. *et al.* Distinct effects of T-bet in TH1 lineage commitment and IFN- $\gamma$  production in CD4 and CD8 T cells. *Science* **295**, 338–342 (2002).
- Cella, M., Otero, K. & Colonna, M. Expansion of human NK-22 cells with IL-7, IL-2, and IL-1 $\beta$  reveals intrinsic functional plasticity. *Proc. Natl Acad. Sci. USA* **107**, 10961–10966 (2010).
- Crellin, N. K. *et al.* Regulation of cytokine secretion in human CD127<sup>+</sup> LT-like innate lymphoid cells by Toll-like receptor 2. *Immunity* **33**, 752–764 (2010).
- Zheng, Y. *et al.* Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nature Med.* **14**, 282–289 (2008).
- Sonnenberg, G. F., Monticelli, L. A., Elloso, M. M., Fouser, L. A. & Artis, D. CD4<sup>+</sup> lymphoid tissue-inducer cells promote innate immunity in the gut. *Immunity* **34**, 122–134 (2011).
- Cash, H. L., Whitham, C. V., Behrendt, C. L. & Hooper, L. V. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* **313**, 1126–1130 (2006).
- Garrett, W. S. *et al.* Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system. *Cell* **131**, 33–45 (2007).
- Awoniyi, M., Miller, S. I., Wilson, C. B., Hajjar, A. M. & Smith, K. D. Homeostatic regulation of *Salmonella*-induced mucosal inflammation and injury by IL-23. *PLoS ONE* **7**, e37311 (2012).
- Klose, C. S., Hoyer, T., Kiss, E. A., Tanriver, Y. & Diefenbach, A. Transcriptional control of innate lymphocyte fate decisions. *Curr. Opin. Immunol.* **24**, 290–296 (2012).
- Hirota, K. *et al.* Fate mapping of IL-17-producing T cells in inflammatory responses. *Nature Immunol.* **12**, 255–263 (2011).
- Bernink, J. H. *et al.* Human type 1 innate lymphoid cells accumulate in inflamed mucosal tissues. *Nature Immunol.* <http://dx.doi.org/10.1038/ni.2534> (in the press).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank G. Häcker for support, the members of the Diefenbach laboratory for valuable discussions, and C. A. Connor for comments on the manuscript. We are grateful to M. Follo, K. Geiger and J. Bodinek-Wersing for cell sorting. The work was supported by the Deutsche Forschungsgemeinschaft (SFB620/A14 to A.D. and TA436/2-1 to Y.T.), the German Federal Ministry of Education and Research (BMBF 01 EO 0803 to A.D. and Y.T.), an EFSD/Lilly grant (to Y.T.), a European Research Council Grant (to A.D.), a CCI Walter-Hitzig-Fellowship (to V.S.), an EMBO Long Term Fellowship (ALTF 508-2011 to A.L.C.) and the FP6 Marie Curie Research Training Network (grant MRTN-CT-2004-005632 (IMDEMI) to A.W.).

**Author Contributions** C.S.N.K. carried out most of the experiments and analysed data, E.A.K. performed infection experiments, histology and immunofluorescence staining, V.S. carried out the analysis of microbial communities, K.E. and N.G. contributed to the infection experiments, T.H. contributed to the analysis of lymphoid organogenesis, Y.d'H. carried out immunofluorescence staining, and A.L.C. and A.W. generated and provided *IL-17 $\gamma$ -cre*<sup>Tg</sup> mice. A.D. and Y.T. designed the study, coordinated the experimental work and analysed data. A.D. directed the research and wrote the manuscript with input from co-authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.D. ([andreas.diefenbach@uniklinik-freiburg.de](mailto:andreas.diefenbach@uniklinik-freiburg.de)).

## METHODS

**Mouse strains.** C57BL/6 and BALB/c mice were purchased from Janvier Laboratories. *Tbx21*<sup>-/-</sup> (ref. 21), *Rorc*( $\gamma$ t)<sup>GFP/+</sup> (ref. 32), *Rorc*( $\gamma$ t)-*cre*<sup>Tg</sup> (ref. 33), *Rosa26R*<sup>YFP/YFP</sup> (ref. 34), *Ahr*<sup>-/-</sup> (ref. 35), *Il12rb2*<sup>-/-</sup> (ref. 36), *Il23p19*<sup>-/-</sup> (ref. 37), *Il17f*-*cre*<sup>Tg</sup> (refs. 38, 39), *Myd88*<sup>fl/fl</sup> (ref. 40) and B6-Ly5.2/Cr (CD45.1<sup>+</sup>) mice were bred in our facilities. *Rag2*<sup>-/-</sup> on a C57BL/6 background and *Rag2*<sup>-/-</sup> *Il2rg*<sup>-/-</sup> mice on a BALB/c background were provided by C. Johner. *Stat4*<sup>-/-</sup> mice<sup>41</sup> were a gift from M. Freudenberg. *Myd88*<sup>-/-</sup> *Trif*<sup>-/-</sup> mice were generously provided by C. Kirschning<sup>42</sup>. Germ-free mice were obtained from K. McCoy. Animals were 6 to 16 weeks of age unless indicated otherwise. All animal experiments were approved and are in accordance with the local animal care committees (Regierungspräsidium Freiburg).

**Isolation of lamina propria leukocytes.** Lamina propria leukocytes were purified as described previously<sup>43</sup>. In brief, small intestine, caecum or colon was removed and cleaned and Peyer's patches were identified and eliminated. After washing with ice-cold PBS, dissociation of epithelial cells was carried out by incubation at 37 °C in Hank's Balanced Salt Solution (HBSS) containing 5 mM EDTA and 10 mM HEPES and vortexing twice. Afterwards, the remaining lamina propria fraction was minced with a scalpel and incubated in HBSS containing Dispase (5 units ml<sup>-1</sup>; BD Biosciences), Collagenase D (0.5 mg ml<sup>-1</sup>; Roche) and DNaseA (0.5 mg ml<sup>-1</sup>; Sigma-Aldrich), vortexed and collected. In some experiments Dispase was not included in the enzyme cocktail. Lymphocytes were purified by Percoll gradient centrifugation (Sigma-Aldrich). Data shown throughout the study are from small intestine of 6- to 16-week-old mice unless stated otherwise.

**Cell sorting, and *in vivo* and *in vitro* differentiation.** Single-cell suspensions of lamina propria lymphocytes were sorted (purity >98%) using a BD FACSaria III cell sorter (BD Biosciences). For adoptive transfer experiments, 5,000 highly purified ILCs (H-2<sup>b</sup>) were injected intravenously in BALB/c *Rag2*<sup>-/-</sup> *Il2rg*<sup>-/-</sup> (H-2<sup>d</sup>) mice for *in vivo* differentiation and mice were killed 3 weeks after transfer. For *in vitro* differentiation, 100 cells were cultured in DMEM supplemented with 10% FCS, 80 mM 2-mercaptoethanol, 10 mM HEPES, 1 mM sodium pyruvate plus non-essential amino acids (Gibco), 80  $\mu$ M 2-mercaptoethanol, 8 mg ml<sup>-1</sup> glutamine, 100 units ml<sup>-1</sup> penicillin, 0.4 mg ml<sup>-1</sup> gentamicin and 100  $\mu$ g ml<sup>-1</sup> streptomycin in 96-well micro-titre plates. SCF, IL-7 and IL-2 (20 ng ml<sup>-1</sup> each) were added to the cultures.

**Generation of bone-marrow chimaeras.** B6-Ly5.2/Cr mice (CD45.1) were lethally  $\gamma$ -irradiated (11 grays) and reconstituted with a 1:1 mixture of bone marrow cells either from *Tbx21*<sup>+/-</sup> mice (CD45.1<sup>+</sup> or CD45.2<sup>+</sup>) or from CD45.1<sup>+</sup> *Tbx21*<sup>+/-</sup> and CD45.2<sup>+</sup> *Tbx21*<sup>-/-</sup> mice. For the first 2 weeks, mice received antibiotic-containing drinking water (2 g l<sup>-1</sup> neomycin sulphate, Gibco; 100 mg l<sup>-1</sup> polymyxin B sulphate, Roth).

**Citrobacter rodentium and Salmonella typhimurium infection.** Infections with *Citrobacter rodentium* and *Salmonella typhimurium* were carried out as described previously<sup>8,12</sup>. *Citrobacter rodentium* ICC169 (a gift from M. Horne) was grown in lysogeny broth medium supplemented with nalidixic acid (30  $\mu$ g ml<sup>-1</sup>) at 37 °C. From an overnight culture, a fresh lysogeny broth medium was inoculated and incubated until an attenuation  $D_{600\text{nm}}$  of 0.2 was reached as determined by an Amersham Ultraspec 2100pro spectrophotometer using a 10-mm quartz cuvette. Under these conditions, an attenuation  $D_{600\text{nm}}$  of 0.2 corresponded to 10<sup>8</sup> bacteria per ml. Bacteria (10<sup>10</sup>) in 200  $\mu$ l PBS were administered to each mouse by oral gavage. Mice were fasted 4 h before oral gavage. Organs and faeces were weighed and homogenized before serial dilution. Bacterial titres were determined after plating the serial dilutions on MacConkey agar plates and incubation overnight at 37 °C. For *Salmonella typhimurium* infection, mice were fasted for 4 h and received 20 mg streptomycin per mouse via oral gavage 24 h before infection. *S. enterica* serovar Typhimurium SL1344 (SB300; a gift from W. D. Hardt) was grown in lysogeny broth medium at 37 °C. An overnight culture was used to inoculate a new bacterial culture that was grown until  $D_{600\text{nm}}$  was 0.8 and corresponded to 10<sup>8</sup> bacteria per ml as confirmed by plating of the inoculum. An inoculum of 10<sup>5</sup> or 10<sup>7</sup> bacteria in 200  $\mu$ l PBS per mouse was taken to infect mice per oral gavage. Mice were fasted 4 h before oral gavage. Depletion of ILCs and natural killer cells has been described previously<sup>7</sup>.

**Histology and pathology score.** Caecal or ileal tissue was cleaned by washing in PBS before fixation in 4% PFA at 4 °C for 2 h. After the fixation process, tissue was washed again in PBS before rehydrating the tissue in 30% (w/v) sucrose solution overnight. Afterwards, the fixed tissue was embedded in OCT compound (Tissue-Tek), snap-frozen in liquid nitrogen and stored at -80 °C. Cryostat sections of 7- $\mu$ m thickness were collected on frosted glass slides and dried at 37 °C for 2 h. Cecums were stained with haematoxylin and eosin (H&E). The following histological parameters were evaluated: leukocyte infiltration, epithelial injury, submucosal oedema and the numbers of goblet cells. Leukocyte infiltration and epithelial injury were scored as absent, mild (1), moderate (2) or severe (3). Submucosal oedema was scored based on the average widths of the submucosa from five

random measurements and the area covered by the oedema. A score of 0 was given for no oedema; a score of 1 was given if the oedema accounted for <50% of the submucosal area and was <100  $\mu$ m wide; a score of 2 was given if the oedema accounted for 50–80% of the submucosal area and was <200  $\mu$ m wide; and a score of 3 was given if the oedema accounted for >80% of the submucosal area and was >200  $\mu$ m wide. The average number of goblet cells was counted in five visual fields (magnification,  $\times 20$ ) and the average number of goblet cells per visual field was calculated. A score of 0 was given for >40 goblet cells per visual field; a score of 1 for 26–40 goblet cells per visual field; a score of 2 for 16–25 goblet cells per visual field; and a score of 3 for 0–15 goblet cells per visual field.

**Quantification of microflora by culture.** Faeces were collected from the small and large intestines and were weighed. Bacteria were quantified by plating serial dilutions on culture plates for aerobic and anaerobic bacteria. Aerobic cultures were incubated for 24–48 h. Anaerobic bacteria were incubated in jars with CO<sub>2</sub> for 48–72 h. After culture, the colonies were counted and single colonies were picked and identified by MALDI-TOF (Bruker).

**Determination of microbial communities by 16S ribosomal DNA quantitative PCR.** Faeces were collected from the small intestine or the colon or rectum segment from co-housed *Tbx21*<sup>-/-</sup> and *Tbx21*<sup>+/-</sup> mice. Bacterial genomic DNA was extracted using the MoBio Powersoil bacterial DNA isolation kit as described previously<sup>43</sup>. The abundance of specific bacterial groups was quantified by real-time PCR using 16S ribosomal DNA gene primers. First, a universal primer was used to determine the total bacterial load in each intestinal segment. Next, quantitative PCR was carried out with group-specific primers to enumerate the amount of bacteria in each of the following major groups: Bacteroidetes (mouse intestinal *Bacteroides*, *Bacteroides* species), Firmicutes (*Lactobacillus* species, segmented filamentous bacteria)<sup>44,45</sup>.

**Antibodies, immunofluorescence staining and cytokine stimulation.** Single-cell suspensions of purified cells were kept on ice and Fc receptors were blocked first with CD16 or CD32 antibody. Lineage-positive cells were excluded using an antibody cocktail specific for CD3, CD8 $\alpha$ , CD19 and Ly6C and Ly6G. For cytokine staining, the indicated cells were incubated in medium supplemented with brefeldin A (10  $\mu$ g ml<sup>-1</sup>) for 4 h at 37 °C. PMA (25 ng ml<sup>-1</sup>) and ionomycin (500 ng ml<sup>-1</sup>) or cytokines (50 ng ml<sup>-1</sup> unless stated otherwise) were added to some wells for stimulation. The following conjugated antibodies were purchased from eBioscience: CCR7 (4B12), CD3 (145-2C11), CD4 (L3T4), CD19 (1D3), CD25 (PC61), CD45.1 (A20), CD127 (A7R34), CD11b (M1/70), CD44 (IM7), CD45.2 (104), CD8 $\alpha$  (53-6.7), CD90.2 (53-2.1), CXCR3 (CXCR3-173), Ly6C and Ly6G (RB6-8C5), H2-K<sup>b</sup> (AF6-88.5), IFN- $\gamma$  (XMG1.2), IL-22 (IH8PWSR), Kit (ACK2), NKp46 (29A1.4), ROR $\gamma$ t (B2D), Sca-1 (D7), T-bet (eBIO4B10) and TNF (MP6-XT22). IL-17A (TC11-18H10), IL-12RB1 (114) and Ki67 (B56) antibodies were ordered from BD Biosciences and CCR6 (29-2L17) antibody from Biolegend. All flow cytometry experiments were acquired using a FACS Canto II flow cytometer and the FACS Diva software (BD Biosciences). For data analysis, FlowJo V8.8 software (TreeStar) was used.

**Quantitative real-time PCR.** Purified cells were resuspended in Trizol (Invitrogen) and stored at -80 °C. RNA was purified by chloroform extraction and RNA concentration was determined using a Nanodrop 8000 device (Thermo Scientific). RNA was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit according to the protocol provided by the manufacturer (Applied Biosystems). Samples were detected on a ABI Prism 7900 sequence detector (Applied Biosystems) using either SYBR Green (Applied Biosystems) reagent and primers as depicted in Supplementary Table 1 or the following TaqMan Gene Expression Assays (Applied Biosystems): *Ahr* (Mm01291777\_m1), *Gzmb* (Mm00442834\_m1), *Ifng* (Mm99999071\_m1), *Il17f* (Mm00521423\_m1), *Kit* (Mm00445212\_m1), *Lta* (Mm00440227\_m1), *Ltb* (Mm00434774\_g1), *Prfl* (Mm00812512\_m1) and *Tbx21* (Mm00450960\_m1). Gene expression was normalized as *n*-fold difference to the gene *Hprt1* (Mm00446968) according to the cycling threshold. Calculation of mRNA levels was performed with SDS2.4 software (Applied Biosystems).

**Primer sequences.** Primer sequences for gene-expression quantification by real-time PCR were determined using PrimerBank, a public resource for PCR primers. The primer sequences were as follows: *Cxcr3*, forward 5'-TACCTTGAGGTTAGTGAACGTC-3', reverse 5'-CGCTCTCGTTTCCCCATAATC-3'; *Reg3g*, forward 5'-TTCCTGTCTCCATGATCAAAA-3', reverse 5'-CATCCACCTCTGTGGGTTC-3'; *Ccr6*, forward 5'-CCTGGGCAACATTATGGTGGT-3', reverse 5'-CAGAACGGTAGGGTGAGGACA-3'; *Il17a*, forward 5'-TTTAACTCCCTTGGCGCAAAA-3', reverse 5'-CTTTCCTCCGCAATTGACAC-3'; *Cd3*, forward 5'-TTCTCTGTACCATGACACTGTC-3', reverse 5'-CGTGGAACTTCCGGCTGTAG-3'; *Ccl4*, forward 5'-TTCCTGCTGTTTCTCTTACACCT-3', reverse 5'-CTGTCTGCCTCTTTTGGTCAG-3'; *FasL*, forward 5'-TCCGTGAGTTCACCAACCAAAA-3', reverse 5'-GGGGGTTCCCTGTAAATGGG-3'; *Il22*, forward 5'-CATGCAGGAGGTGGTACCTT-3', reverse 5'-CAGACGCAAGCA

TTTCTCAG-3'; Il23, forward 5'-ATGCTGGATTGCAGAGCAGTA-3', reverse 5'-ACGGGGCACATTATTTTCTAGTCT-3'; *Light*, forward 5'-CAGGCCCTACAGACAACAC-3', reverse 5'-ACTCGTCTCCAGGAAC-3'; *Rorc*( $\gamma$ t), forward 5'-TGCTCTGGGCTACCCTACTG-3', reverse 5'-GTAGTGGATCCCA GATGACTTGT-3'; *Tnf*, forward 5'-CCCTCACACTCAGATCATCTTCT-3', reverse 5'-GCTACGACGTGGGCTACAG-3'.

**Statistical analysis.** Statistical significance of the data sets was determined by performing two-tailed Student's *t*-tests with 95% confidence intervals. If equal variances between the groups could not be assumed, Welch's correction of the *t*-test was applied. Tests for statistical significance were performed with Graph Pad Prism V4 software. Data were calculated as mean  $\pm$  standard deviation (s.d.). To improve clarity, only mean + s.d. is shown in the figures.

32. Eberl, G. *et al.* An essential function for the nuclear receptor ROR $\gamma$ t in the generation of fetal lymphoid tissue inducer cells. *Nature Immunol.* **5**, 64–73 (2004).
33. Eberl, G. & Littman, D. R. Thymic origin of intestinal  $\alpha\beta$  T cells revealed by fate mapping of ROR $\gamma$ t<sup>+</sup> cells. *Science* **305**, 248–251 (2004).
34. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of *EYFP* and *ECFP* into the *ROSA26* locus. *BMC Dev. Biol.* **1**, 4 (2001).
35. Schmidt, J. V., Su, G. H., Reddy, J. K., Simon, M. C. & Bradfield, C. A. Characterization of a murine Ahr null allele: involvement of the Ah receptor in hepatic growth and development. *Proc. Natl Acad. Sci. USA* **93**, 6731–6736 (1996).
36. Wu, C. *et al.* IL-12 receptor  $\beta$ 2 (IL-12R  $\beta$ 2)-deficient mice are defective in IL-12-mediated signaling despite the presence of high affinity IL-12 binding sites. *J. Immunol.* **165**, 6221–6228 (2000).
37. Ghilardi, N. *et al.* Compromised humoral and delayed-type hypersensitivity responses in IL-23-deficient mice. *J. Immunol.* **172**, 2827–2833 (2004).
38. Kurschus, F. C. *et al.* Genetic proof for the transient nature of the Th17 phenotype. *Eur. J. Immunol.* **40**, 3336–3346 (2010).
39. Croxford, A. L., Kurschus, F. C. & Waisman, A. Cutting edge: an IL-17F-Cre<sup>EYFP</sup> reporter mouse allows fate mapping of Th17 cells. *J. Immunol.* **182**, 1237–1241 (2009).
40. Hou, B., Reizis, B. & DeFranco, A. L. Toll-like receptors activate innate and adaptive immunity by using dendritic cell-intrinsic and -extrinsic mechanisms. *Immunity* **29**, 272–282 (2008).
41. Kaplan, M. H., Sun, Y. L., Hoey, T. & Grusby, M. J. Impaired IL-12 responses and enhanced development of Th2 cells in Stat4-deficient mice. *Nature* **382**, 174–177 (1996).
42. Conrad, M. L. *et al.* Maternal TLR signaling is required for prenatal asthma protection by the nonpathogenic microbe *Acinetobacter lwoffii* F78. *J. Exp. Med.* **206**, 2869–2877 (2009).
43. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
44. Salzman, N. H. *et al.* Analysis of 16S libraries of mouse gastrointestinal microflora reveals a large new group of mouse intestinal bacteria. *Microbiology* **148**, 3651–3660 (2002).
45. Salzman, N. H. *et al.* Enteric defensins are essential regulators of intestinal microbial ecology. *Nature Immunol.* **11**, 76–82 (2010).



# A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis

Paola Picotti<sup>1,2\*</sup>, Mathieu Clément-Ziza<sup>3\*</sup>, Henry Lam<sup>4\*</sup>, David S. Campbell<sup>5</sup>, Alexander Schmidt<sup>6</sup>, Eric W. Deutsch<sup>5</sup>, Hannes Röst<sup>1</sup>, Zhi Sun<sup>5</sup>, Oliver Rinner<sup>1,7</sup>, Lukas Reiter<sup>1,7</sup>, Qin Shen<sup>1,8</sup>, Jacob J. Michaelson<sup>3</sup>, Andreas Frei<sup>1</sup>, Simon Alberti<sup>9</sup>, Ulrike Kusebauch<sup>5</sup>, Bernd Wollscheid<sup>1</sup>, Robert L. Moritz<sup>5</sup>, Andreas Beyer<sup>3</sup> & Ruedi Aebersold<sup>1,10,11</sup>

Experience from different fields of life sciences suggests that accessible, complete reference maps of the components of the system under study are highly beneficial research tools. Examples of such maps include libraries of the spectroscopic properties of molecules, or databases of drug structures in analytical or forensic chemistry. Such maps, and methods to navigate them, constitute reliable assays to probe any sample for the presence and amount of molecules contained in the map. So far, attempts to generate such maps for any proteome have failed to reach complete proteome coverage<sup>1–3</sup>. Here we use a strategy based on high-throughput peptide synthesis and mass spectrometry to generate an almost complete reference map (97% of the genome-predicted proteins) of the *Saccharomyces cerevisiae* proteome. We generated two versions of this mass-spectrometric map, one supporting discovery-driven (shotgun)<sup>3,4</sup> and the other supporting hypothesis-driven (targeted)<sup>5,6</sup> proteomic measurements. Together, the two versions of the map constitute a complete set of proteomic assays to support most studies performed with contemporary proteomic technologies. To show the utility of the maps, we applied them to a protein quantitative trait locus (QTL) analysis<sup>7</sup>, which requires precise measurement of the same set of peptides over a large number of samples. Protein measurements over 78 *S. cerevisiae* strains revealed a complex relationship between independent genetic loci, influencing the levels of related proteins. Our results suggest that selective pressure favours the acquisition of sets of polymorphisms that adapt protein levels but also maintain the stoichiometry of functionally related pathway members.

In proteomics, the generation of reference maps covering a complete proteome has been attempted in two ways. The first is based on the development of immunoassays to detect target proteins and is exemplified for the human proteome by the Protein Atlas project<sup>8</sup>. The second approach is in-depth mapping of a proteome through the collection of fragment ion spectra from multiple mass-spectrometry-based shotgun proteomic experiments<sup>3,4,9</sup>. Such reference spectra can be used in discovery-driven experiments to analyse acquired fragment ion spectra using spectral matching<sup>10–12</sup>, or in targeted measurements, to specifically monitor proteins of interest by selected reaction monitoring (SRM)<sup>5,6,13</sup>. So far, both approaches have failed to reach complete proteome coverage; in the case of yeast<sup>1–3</sup> and other microbes<sup>14,15</sup>, saturation is apparent when about two-thirds of all proteins that would be predicted based on the genome are covered, and this coverage is much lower for other proteomes, including the human proteome<sup>1</sup>.

We defined the yeast proteome as the collection of 6,607 protein sequences predicted based on the genome of yeast, each one associated with an open reading frame (ORF) in the *Saccharomyces* Genome

Database (SGD, <http://www.yeastgenome.org>). First, we classified yeast proteins based on their detectability, using a large repository of proteomic data, PeptideAtlas<sup>1,3</sup>, and the largest data set of antibody-based protein measurements in yeast, which quantified a common tag engineered into each ORF (Supplementary Fig. 1a)<sup>2</sup>. The coverage of yeast ORFs was below two-thirds of the ORFeome for each of the two orthogonal data sets, suggesting that the proteome of yeast grown under standard laboratory conditions has been exhaustively mapped out by automated peptide sequencing or by antibody-based detection, and the two orthogonal data sets showed a high degree of overlap. Next, for each protein from PeptideAtlas we selected an optimal set of up to eight peptides with favourable mass-spectrometry properties and unique occurrence within the compiled protein sequence database (proteotypic peptides, PTPs)<sup>16</sup>. We predicted PTPs for proteins for which no empirical data were available (Supplementary Fig. 1b). For proteins that were not included in PeptideAtlas and that were not detected using the antibody-based methods, we selected at least two peptides with an isoelectric point below 4.5, if available, to maximize the probability of detecting the corresponding proteins if the peptide samples were first fractionated by off-gel electrophoresis<sup>5</sup>. Approximately 200 proteins remained refractory to these selection criteria, including proteins that do not generate any suitable tryptic peptide for mass-spectrometry analysis (Supplementary Fig. 1c). The final peptide set, comprising approximately 28,000 peptides, was synthesized on a small scale to assemble a peptide library representing 97% of the predicted yeast proteome (Fig. 1).

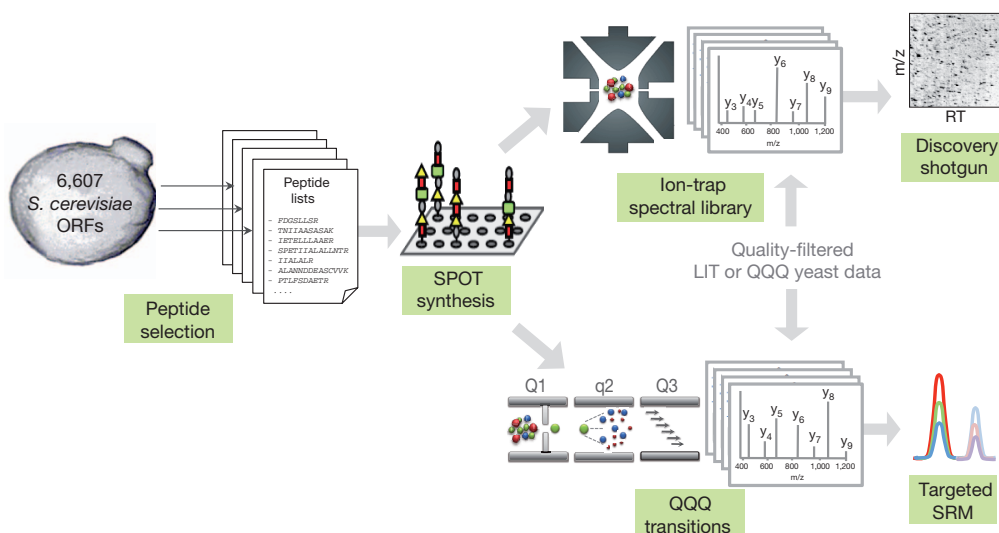
We next used these peptides to generate two reference spectral libraries, each one supporting a commonly used proteomic method. We analysed the peptide set on a linear ion trap (LIT)-type instrument (LIT–Orbitrap hybrid) to generate reference fragment ion spectra for spectral matching of data acquired in discovery experiments. We used a QQQ-type mass spectrometer (QTRAP hybrid) operated in the SRM-triggered tandem mass spectrometry (MS/MS) mode with fragmentation in the second quadrupole<sup>17</sup> to generate fragment ion spectra for the extraction of optimal SRM coordinates for targeted measurement of specific proteins (Fig. 1). The spectra acquired with each instrument were assigned to peptide sequences by sequence-database searching and the assignments were filtered to a false discovery rate (FDR) at a peptide-spectrum match level of less than 0.1% and at a peptide level of less than 0.5%, based on decoy counts.

To maximize proteome coverage, we combined our LIT data with quality-filtered LIT data from yeast extracts that had been submitted to PeptideAtlas<sup>1</sup> and consensus spectra from the National Institute of Standards and Technology (NIST) yeast ion-trap spectral library (<http://peptide.nist.gov>; build 19 October 2009). The QQQ data were

<sup>1</sup>Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich CH-8093, Switzerland. <sup>2</sup>Department of Biology, Institute of Biochemistry, ETH Zurich, Zurich CH-8093, Switzerland.

<sup>3</sup>Biotechnology Centre, Technische Universität Dresden, D-01069 Dresden, Germany. <sup>4</sup>Department of Chemical and Biomolecular Engineering and Division of Biomedical Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. <sup>5</sup>Institute for Systems Biology, Seattle, Washington 98103, USA. <sup>6</sup>Proteomics Core Facility Biozentrum, University of Basel, Klingelbergstrasse 50/70, CH-4056 Basel, Switzerland. <sup>7</sup>Biognosys AG, Zurich CH-8093, Switzerland. <sup>8</sup>Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 200031 Shanghai, China. <sup>9</sup>Max Planck Institute of Molecular Cell Biology and Genetics, D-01307 Dresden, Germany. <sup>10</sup>Competence Center for Systems Physiology and Metabolic Diseases, Zurich CH-8093, Switzerland. <sup>11</sup>Faculty of Science, University of Zurich, Zurich CH-8057, Switzerland.

\*These authors contributed equally to this work.



**Figure 1 | Generation of a reference mass-spectrometric map for the yeast proteome.** Schematic of the sequential steps of map generation: peptide selection based on the known ORFeome, small-scale peptide synthesis, QQQ- and LIT-type mass-spectrometry measurements, pooling of the generated spectra to existing data sets and application of the libraries to SRM measurements or spectral searches of yeast shotgun data sets.  $m/z$ , mass-to-

charge ratio; Q1, first quadrupole of a triple-quadrupole mass spectrometer; Q2, second quadrupole of a triple-quadrupole mass spectrometer (quadrupolar fragmentation cell); Q3, third quadrupole of a triple-quadrupole mass spectrometer; RT, retention time. On the SPOT synthesis membrane, different coloured shapes indicate different amino acids. Specific shapes and colours are not assigned to specific amino acids.

combined<sup>11</sup> with the data set acquired from yeast proteins from the original MRMatlas<sup>17</sup>. For the assigned peptides, high-quality, de-noised consensus spectra were compiled<sup>11</sup>. The final LIT and QQQ spectral libraries contained consensus spectra for 100,815 and 28,216 peptide sequences, respectively, both covering 97% of the 6,607 sequences in the yeast proteome (Fig. 2a). For each peptide we then extracted from the QQQ spectral library the charge state (or states), information of the prominent fragment ions, including their masses, charges, relative intensities and chromatographic elution times, which collectively constitute a peptide SRM assay.

Our peptide selection criteria resulted in a peptide set that preferentially contained peptides of intermediate hydrophobicity (Fig. 2b). The approximately 1,630 synthetic peptides that could not be detected showed a bias towards extreme calculated<sup>18</sup> hydrophobicity values (Fig. 2b), indicating that very hydrophilic or very hydrophobic peptides are not well suited for chemical synthesis and/or liquid chromatography–mass spectrometry (LC–MS) analysis. Spectra acquired from synthetic peptides were indistinguishable from those acquired from the corresponding natural sources<sup>17</sup> (Supplementary Fig. 2).

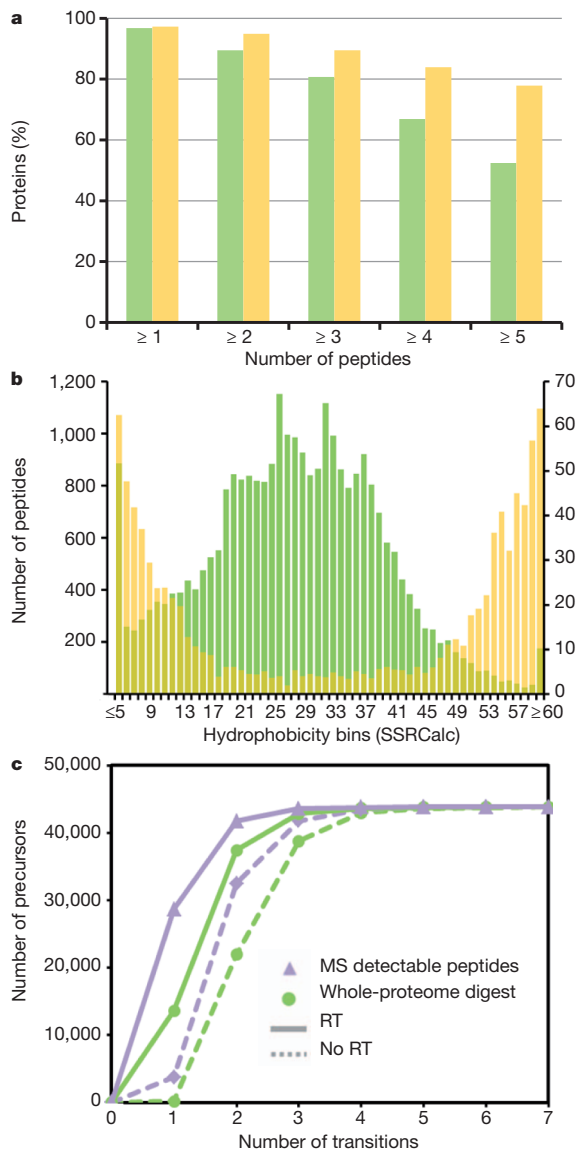
The LIT and QQQ data sets were compiled into two ‘builds’ in PeptideAtlas<sup>1</sup>, and can be downloaded or browsed interactively (Supplementary Discussion, Supplementary Figs 3–7 and <http://www.srmatlas.org/yeast/>).

We assessed the performance of the LIT library using a total yeast tryptic digest, analysed by data-dependent LC–MS/MS (data-dependent acquisition (DDA) mode). Spectral-library searching against the LIT library identified 1,617 unique proteins (Supplementary Fig. 8), compared to 1,529 proteins identified by conventional database searching<sup>19</sup>, at the same protein FDR of 1%. Therefore, restricting the analysis to our chosen PTPs did not diminish the number of proteins identified. We next analysed the yeast sample multiple times, using distinct inclusion lists containing all precursor ions corresponding to the five top-ranking peptides per protein in the library. Spectral searching identified 2,509 unique proteins (Supplementary Figs 8 and 9). We then re-analysed the same sample using a single, reduced inclusion list containing only the peptide ions identified in the prior step (Supplementary Data set). Spectral matching against the LIT library identified 1,987 proteins (Supplementary Figs 8–10). This workflow resulted in approximately 30% more identifications than those achieved by conventional DDA followed by

sequence-database searching. The spectral library search of this single file of about 34,000 MS/MS spectra took 5 min on one processor (approximately 0.01 s per spectrum), which implies that conducting on-the-fly data analysis is feasible in discovery-based proteomic experiments.

We next analysed the specificity of our SRM transitions in terms of uniqueness against two backgrounds of different complexity (Fig. 2c), using the SRMCollider tool<sup>20</sup> (Supplementary Discussion). Of the peptide precursors, 97.8% and 88.5% (Fig. 2c and <http://www.srmatlas.org/yeast/>) were predicted to be uniquely detected using the three highest transitions in our library, with and without time-scheduled acquisition, respectively, in the high complexity background. As simulations do not necessarily reflect the full complexity of a biological sample, it is advisable to measure four to five transitions per peptide and to use empirical relative intensities of fragment ions and peptide elution times as constraints to increase the assay specificity.

To demonstrate the utility of the two libraries in discovery and targeted proteomic experiments, we applied them to a protein-based quantitative trait locus (QTL) analysis in *S. cerevisiae*. Protein QTL studies aim to correlate protein abundance with genetic variation, and thus critically rely on the ability to measure protein concentrations precisely throughout large numbers of samples. Previous protein QTL studies<sup>7</sup> suffered from an inconsistent detection of peptides across samples (when a given peptide is present, it is not consistently detected) and a bias towards the detection of abundant proteins (Supplementary Discussion). To overcome these limitations we applied a two-step workflow based on our spectral libraries to a genetically diverse population of 78 yeast strains obtained by crossing a wild isolate (RM11-1a) and a strain isogenic to the standard S288c laboratory strain (BY4716)<sup>21</sup>. To identify proteins whose cellular concentrations are probably affected by QTLs in this cross, we carried out a discovery proteomic experiment on the two parental yeast strains and a subset of 16 segregants out of the total 78 strains that we selected, with the aim to maximize the genetic diversity between them (Supplementary Discussion and Supplementary Fig. 11). Using the single inclusion list described above and spectral matching against the LIT library, we identified approximately 2,500 uniquely mapped proteins at 1% FDR and quantified the abundance of the corresponding peptides throughout the different samples using a label-free approach. We excluded peptides with sequence differences between the RM11-1a and BY4716 backgrounds and proteins for which no high-quality



**Figure 2 | Composition and use of the spectral libraries.** **a**, Number of peptides per protein in the SRM assay (green) or ion-trap (yellow) library. The percentage is relative to the total of 6,607 theoretical *S. cerevisiae* proteins. **b**, Hydrophobicity distribution for all selected peptides (green, left axis) and for those that could not be detected with our QQQ workflow (yellow, right axis). Hydrophobicity is expressed as sequence-specific retention calculator (SSRCalc)<sup>18</sup> score bins. A low SSRCalc score indicates low hydrophobicity. **c**, Minimal number of transitions predicted to uniquely identify precursor ions in the QQQ library. A cumulative plot of the minimal number of transitions forming an assay (unique ion signature<sup>20</sup>) is shown for all the precursors in the QQQ library. Transitions were selected with decreasing intensity. Uniqueness for a set of transitions was established if no other peptide precursor produced transitions that contained all of the query transitions, using a tolerance of  $\pm 0.35$  Thomson (Th) for both precursor and fragment ions. For time-scheduled SRM acquisition, only co-eluting peptides were considered, based on predicted retention times<sup>18</sup> and a tolerance of  $\pm 2.5$  min. The analysis is shown for two different backgrounds: all yeast peptides potentially detectable by mass spectrometry, as derived from the yeast PeptideAtlas (purple), and all theoretical tryptic peptides in the yeast proteome (green). RT and no RT indicate whether time-scheduled SRM acquisition is considered.

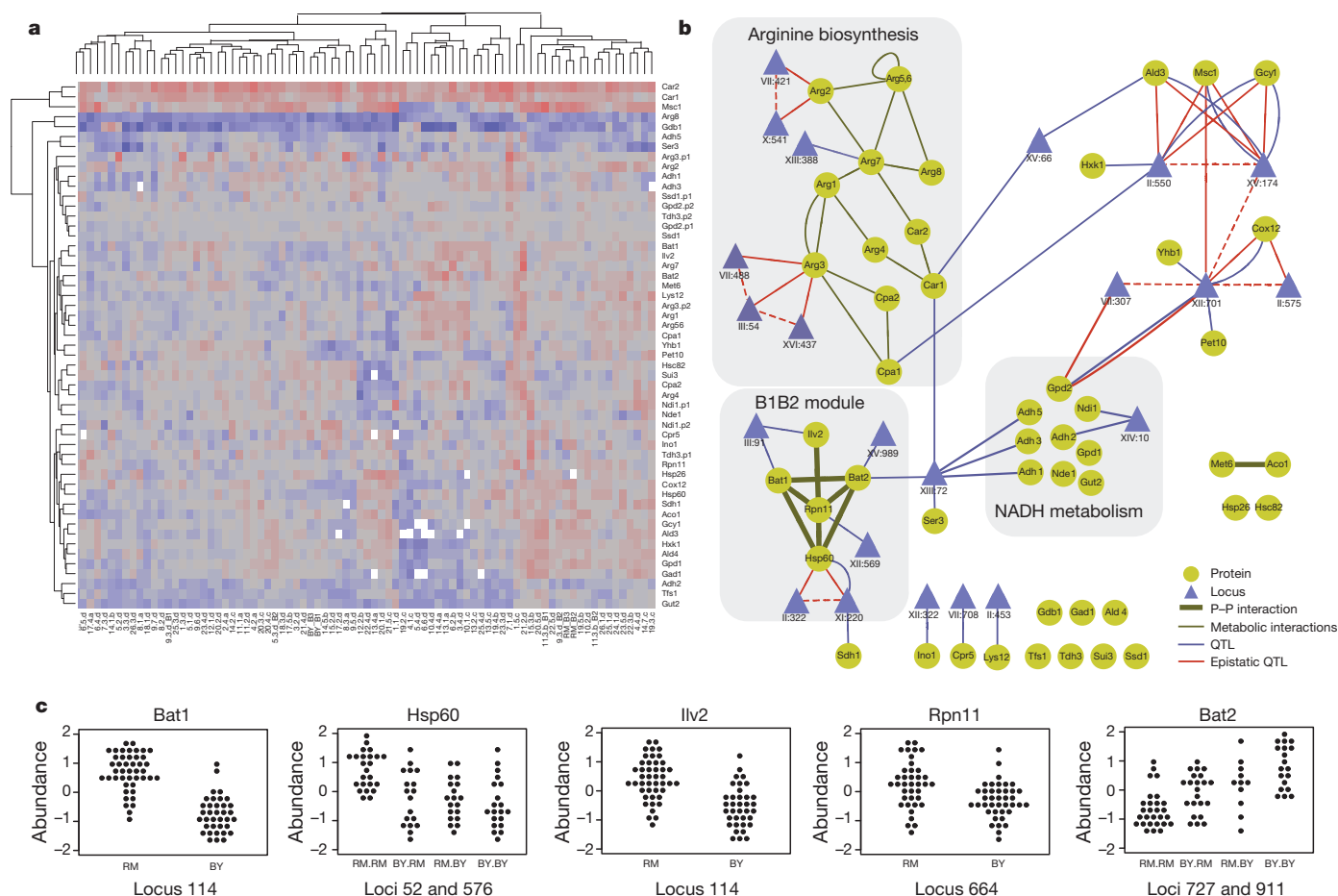
ion features could be detected. We ranked the remaining approximately 2,100 proteins based on the variability of their concentrations across the sample strains, and corrected for the number of peptides that were measured per protein (Supplementary Fig. 12). To identify cellular processes and pathways that were particularly affected by genetic variation, we analysed the 150 proteins showing the

highest variability for gene-ontology-based functional enrichment, metabolic-pathway enrichment, and protein module over-representation tests using protein-interaction data (Supplementary Tables 1–3). Proteins involved in NADH oxidation, arginine or ornithine biosynthesis, and amino acid metabolism were significantly ( $P < 0.01$ ) enriched among the most variable proteins. We selected a set of 48 proteins that are members of the most highly enriched pathways and subnetworks. These proteins covered most levels of cellular abundances (Supplementary Fig. 13) and included a protein from one of the targeted pathways (Arg3) that was not detected in the discovery phase (Supplementary Table 4). We next used SRM assays (Supplementary Information) from the QQQ map for the targeted quantification of the 48 proteins throughout 76 segregants and both parental strains. The SRM-based quantification resulted in a highly consistent and comprehensive data set (Fig. 3a) and enabled the precise determination of inter-sample variations of protein abundances.

Epistatic interactions between genes are important factors contributing to the variation of complex traits and are thought to partly explain ‘missing heritability’ observed in traditional association studies<sup>22</sup>. However, detecting epistasis in QTL studies is notoriously difficult, mostly owing to lack of sufficient statistical power. We reasoned that because of the precision of SRM data and the relatively large number of samples it should be possible to detect epistasis. We therefore extended a machine-learning-based QTL mapping method that we developed previously<sup>23,24</sup> to also report epistatic interactions between pairs of loci affecting a common protein (Supplementary Figs 14–17). Application of this framework to our protein QTL data identified 32 protein QTLs involving single loci ( $FDR < 0.15$ ) and 10 pairs of epistatic protein QTLs ( $FDR < 0.2$ ) (Fig. 3b). In total, 28 of the 48 proteins were under the control of at least one protein QTL and 23 distinct genomic regions were involved. Thus, protein concentrations are strongly affected by natural genetic variation and epistatic interactions between loci affecting protein levels are a common phenomenon. A protein module consisting of Bat1, Bat2, Rpn11, Hsp60 and Ilv2 (which we termed the ‘B1B2 module’) caught our attention, because all of its components were involved in at least one significant protein QTL (Fig. 3b). The module functionally relates to protein turnover and amino acid metabolism, and is physically connected to mitochondria (Supplementary Discussion). Interestingly, six different genomic regions contained polymorphisms independently affecting the levels of the different proteins in the module. Detailed analysis revealed that segregant strains carrying the BY4716 alleles at the respective loci expressed consistently lower levels of all proteins that are part of this module, with Bat2 being the only exception (Fig. 3c). Importantly, Bat2 favours the reverse metabolic reaction catalysed by Bat1 (catabolism and anabolism of branched-chain amino acids, respectively, Supplementary Discussion)<sup>25</sup>. Thus, the two parental strains have acquired a set of independent genetic variations altering the abundance of B1B2 proteins such that the pathway activity consistently changes in one direction. Our data contained a second example of such coordinated acquisition of independent polymorphisms affecting the regulation of alcohol dehydrogenases (NADH module, Fig. 3b): Adh1, Adh3 and Adh5, which are upregulated in the RM11-1a background, can catalyse the last reaction of the ethanol production pathway, whereas Adh2, which is linked to another locus and down-regulated in the RM11-1a background, catalyses the reverse reaction preferentially (Supplementary Fig. 18)<sup>26,27</sup>.

We have described here a mass-spectrometric map for the near-complete proteome of yeast, to support both discovery-driven, and targeted proteomics experiments. The LIT library can be used for spectral matching of shotgun data, thus exploiting the previously acknowledged benefits of spectral matching, such as speed, confidence or an increase in the number of identifications. A limitation of the library is that it does not include protein post-translational modifications. However, in principle, the same approach can also be applied to generate spectral libraries for modified peptides and our libraries can





**Figure 3 | Quantitative trait analysis from the targeted proteomic data set.** **a**, Two-way cluster analysis of summarized protein abundances measured by SRM in 82 samples (78 strains) of the cross between RM11-1a (RM) and BY4716 (BY). Columns are clustered according to the samples, and rows are clustered according to the proteomic traits. Abundance levels are colour coded in a blue–red gradient (blue, low abundance; red, high abundance; white, missing data). The completeness of the data set reaches 99.5% even though the summarization procedure was conservative and generated additional missing values. **b**, Network representation of protein abundance QTL for the 48 targeted proteins. Locus positions are indicated by the chromosome name followed by the genomic position of the centre of the locus in kilobases. Physical protein–protein interactions (P–P interaction) were obtained from BioGRID (<http://thebiogrid.org>). Metabolic interactions were manually reconstructed

from BioCyc (<http://biocyc.org/>). Epistatic protein QTL always link two loci (connected by a dashed line) that are in epistasis with respect to a protein abundance trait (connected by a solid line). **c**, Protein abundances for the B1B2 module. For each protein, abundances are shown for groups of strains separated according to their genotype at the respective protein QTL. A pair of interacting loci was linked to the variations of Hsp60. The epistatic interaction is clearly visible when the RM11-1a allele is inherited at both loci. In contrast, the effects of the two loci linked to Bat2 are additive. Notably, the directionality of regulation is shared by all components of the module except Bat2 (overexpression when the RM11-1a alleles are inherited). Bat1 and Bat2 are paralogs catalysing the same metabolic reaction in opposite directions. RM.RM, BY.RM, RM.BY and BY.BY indicate allele combinations at two interacting loci.

be iteratively expanded and enriched to improve their coverage and quality. The library of SRM coordinates substantially expands the capabilities of SRM-based targeted proteomic experiments and accelerates their implementation. The availability of QQQ fragment-ion relative intensities and peptide elution times can be used for the automated scoring and statistical evaluation of large-scale SRM data sets with respect to their FDR<sup>28</sup>. A limitation of our SRM assay library is that we do not experimentally determine the sensitivity and specificity of each assay. These properties are sample- and platform-dependent (for example, chromatography, resolution and tuning) and should therefore be determined locally, for any particular sample. In addition, as our reference spectra support the selection of multiple intense transitions for each peptide, one or more can be discarded if found to be locally unspecific. For example, although the peptides that differed between the parental backgrounds were discarded in our protein QTL study, 90% of the SRM library remained usable despite 1% divergence between the reference and the RM11-1a genomes.

The libraries presented here are also a useful blueprint for studying peptide-fragmentation properties in quadrupoles and ion-trap mass

spectrometry and will support the development of new acquisition methods relying on the knowledge of peptide fragmentation patterns. A first example of such a data-independent acquisition method, SWATH (sequential windowed acquisition of all theoretical fragment-ion spectra) mass spectrometry, has been described already<sup>29</sup>.

The application of our libraries to yeast protein QTL analysis resulted in the precise measurement of proteins spanning a broad range of abundances (Supplementary Figs 13 and 19) throughout a large number of samples. This enabled the detection of novel protein QTLs and epistatic interactions, and led to the identification of two cases of co-inheritances of several independent genetic variations that influence the abundance of related proteins in a biologically coherent manner. A deeper analysis of these cases together with previous findings<sup>30</sup> suggests that they are examples for the adaptive evolution of protein-level regulation (Supplementary Discussion).

The proteome map generated here, the publicly accessible tools to navigate it, and the data acquisition and processing strategies that are enabled by such resources expand the capabilities of current proteomics experiments and substantially improve their performance features.

## METHODS SUMMARY

We selected a set of peptides for each of the 6,607 proteins in the *S. cerevisiae* proteome based on existing proteomic data sets or computational prediction. We synthesized the peptide set using SPOT synthesis and analysed the crude peptides in batches on both a QQQ and a LIT mass spectrometer. We pooled the generated spectra with spectra extracted from public repositories. The resulting QQQ and LIT spectral libraries were made publicly available through the PeptideAtlas interface (<http://www.srmatlas.org/yeast/>; Supplementary Methods). We used the libraries to analyse a collection of yeast strains grown in glucose-based medium. Proteins were identified and quantified using a shotgun and a targeted proteomic approach, both guided by the mass spectrometric coordinates contained in the library. Protein QTL mapping was carried out using a modified Random Forest machine learning method that also reports pairs of epistatically interacting loci. FDRs of protein QTLs were obtained by permuting the phenotype vectors 25,000 times and thereby retrieving an individual background score distribution for each genetic marker.

Full methods and supplementary material accompany this paper.

Received 3 February 2011; accepted 30 November 2012.

Published online 20 January 2013.

- Deutsch, E. W. The PeptideAtlas Project. *Methods Mol. Biol.* **604**, 285–296 (2010).
- Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
- de Godoy, L. M. *et al.* Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251–1254 (2008).
- Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
- Picotti, P. *et al.* Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* **138**, 795–806 (2009).
- Anderson, L. & Hunter, C. L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* **5**, 573–588 (2006).
- Foss, E. J. *et al.* Genetic basis of proteome variation in yeast. *Nature Genet.* **39**, 1369–1375 (2007).
- Taussig, M. J. *et al.* ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nature Methods* **4**, 13–17 (2007); erratum **4**, 187 (2007).
- Deutsch, E. W., Lam, H. & Aebersold, R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* **9**, 429–434 (2008).
- Lam, H. & Aebersold, R. Spectral library searching for peptide identification via tandem MS. *Methods Mol. Biol.* **604**, 95–103 (2010).
- Lam, H. *et al.* Building consensus spectral libraries for peptide identification in proteomics. *Nature Methods* **5**, 873–875 (2008).
- Craig, R. *et al.* Using annotated peptide mass spectrum libraries for protein identification. *J. Proteome Res.* **5**, 1843–1849 (2006).
- Lange, V. *et al.* Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4**, 222 (2008).
- Lange, V. *et al.* Targeted quantitative analysis of *Streptococcus pyogenes* virulence factors by multiple reaction monitoring. *Mol. Cell. Proteomics* **7**, 1489–1500 (2008).
- Ahrens, C. H. *et al.* Generating and navigating proteome maps using mass spectrometry. *Nature Rev. Mol. Cell Biol.* **11**, 789–801 (2010).
- Mallick, P. *et al.* Computational prediction of proteotypic peptides for quantitative proteomics. *Nature Biotechnol.* **25**, 125–131 (2007).
- Picotti, P. *et al.* A database of mass spectrometric assays for the yeast proteome. *Nature Methods* **5**, 913–914 (2008).
- Krokhin, O. V. Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-Å pore size C18 sorbents. *Anal. Chem.* **78**, 7785–7795 (2006).
- Craig, R. & Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467 (2004).
- Röst, H., Malmström, L. & Aebersold, R. A computational tool to detect and avoid redundancy in selected reaction monitoring. *Mol. Cell. Proteomics* **11**, 540–549 (2012).
- Brem, R. B. *et al.* Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
- Zuk, O. *et al.* The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl Acad. Sci. USA* **109**, 1193–1198 (2012).
- Michaelson, J. J. *et al.* Data-driven assessment of eQTL mapping methods. *BMC Genomics* **11**, 502 (2010).
- Ackermann, M., Clément-Ziza, M., Michaelson, J. J. & Beyer, A. Teamwork: improved eQTL mapping using combinations of machine learning methods. *PLoS ONE* **7**, e40916 (2012).
- Colón, M. *et al.* *Saccharomyces cerevisiae* Bat1 and Bat2 aminotransferases have functionally diverged from the ancestral-like *Kluyveromyces lactis* orthologous enzyme. *PLoS ONE* **6**, e16099 (2011).
- Young, E. T. & Pilgrim, D. Isolation and DNA sequence of ADH3, a nuclear gene encoding the mitochondrial isozyme of alcohol dehydrogenase in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **5**, 3024–3034 (1985).
- de Smidt, O., du Preez, J. C. & Albertyn, J. Molecular and physiological aspects of alcohol dehydrogenases in the ethanol metabolism of *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **12**, 33–47 (2012).
- Reiter, L. *et al.* mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nature Methods* **8**, 430–435 (2011).
- Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, 0111–016717 (2012).
- Fraser, H. B., Moses, A. M. & Schadt, E. E. Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proc. Natl Acad. Sci. USA* **107**, 2977–2982 (2010).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** This project has been funded in part by ETH Zurich, the Swiss National Science Foundation (3100A0-107679), the National Heart, Lung and Blood Institute, National Institutes of Health (N01-HV-28179), the National Science Foundation MRI (grant 0923536), the Luxembourg Centre for Systems Biomedicine and the University of Luxembourg, and by SystemsX.ch, the Swiss initiative for systems biology. P.P. is supported by a Foerderungsforschung grant from the Swiss National Science Foundation (PP00P3\_133670), by a European Union Seventh Framework Program Reintegration grant (FP7-PEOPLE-2010-RG-277147) and by a Promedica Stiftung (2-70669-11). H. L. is supported by the University Grant Council of the Hong Kong Special Administrative Region Government, China (HKUST DAG08/09.EG02). A.B. is supported by the Klaus Tschira Foundation and by a European Union FP7 HEALTH grant (HEALTH-F4-2008-223539). R.A. is supported by the European Research Council (ERC-2008-AdG 233226) and by SystemsX.ch, the Swiss Initiative for Systems Biology.

**Author Contributions** P.P. and M.C.-Z. carried out the experiments; M.C.-Z., H.L., E.W.D., O.R., L.R., P.P., J.J.M., A.B. and R.A. conceived the data analysis pipeline; P.P., M.C.-Z., H.L., D.S.C., A.S., E.W.D., H.R., Z.S., O.R., L.R., J.J.M. and Q.S. analysed the data; A.S., A.F., Q.S. and U.K. performed mass-spectrometry measurements; P.P., A.B., M.C.-Z., S.A. and R.A. designed the experiments; P.P., M.C.-Z., A.B., H.L., H.R., S.A. and R.A. wrote the manuscript; B.W. and R.L.M. supervised part of the project; and R.A., A.B. and P.P. supervised the project.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.A. ([aebersold@imsb.biol.ethz.ch](mailto:aebersold@imsb.biol.ethz.ch)), A.B. ([andreas.beyer@biotec.tu-dresden.de](mailto:andreas.beyer@biotec.tu-dresden.de)) or P.P. ([paola.picotti@bc.biol.ethz.ch](mailto:paola.picotti@bc.biol.ethz.ch)).

# CAREERS

**UNITED STATES** Intelligence agency will need geospatial expertise **p.273**

**EUROPEAN UNION** University rankings with broad criteria aim for transparency **p.273**

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

ADAPTED FROM: ALBERTO MASNOVO/SHUTTERSTOCK



## SCIENCE COMMUNICATION

# From page to screen

*As print dwindles and online platforms grow, job prospects are changing for science writers.*

BY KENDALL POWELL

**L**ena Groeger is not a conventional science journalist. She develops applications for ProPublica, a non-profit online news organization based in New York. Using software code and data sets, she builds searchable databases and interactive graphics to

accompany investigative-news articles. Ten years ago, such a job did not exist — and an online journalism position might have earned funny looks and consoling remarks from colleagues. Back then, web-based journalism was deemed risky and second-rate compared with ‘real’ journalism at newspapers or magazines. Now, those conventional jobs are in decline,

and Groeger’s former classmates from the Science, Health and Environmental Reporting Program (SHERP) at New York University congratulate her on her position.

Internet-focused careers are popping up in many realms of science communication. The field has gone through a similar revolution before, during the Internet boom of the 1990s — but that ended in a disastrous burst bubble. The current online-news trend seems to have more staying power, thanks to mature business models and readers more inclined to spend time and money online. At the same time, conventional outlets for science journalism are on the wane: last month, the *Columbia Journalism Review* reported that the number of US newspapers with weekly science sections had dropped from 95 in 1989 to 19 in 2012. The field’s future remains in flux.

The past few years have seen dramatic shifts in how the public consumes news. “Everybody using the web becomes their own editor, putting their own ‘newspaper’ together every day,” says Rick Borchelt, director of news and public affairs at the US National Cancer Institute in Bethesda, Maryland. He adds that readers are “curating their own streams” of information through blog subscriptions, web alerts or skimming favourite sites.

This has driven a shift towards a direct-to-consumer model in science journalism, communications and public relations. Rather than relying on conventional publications to get the word out, websites such as LiveScience.com take science news straight to readers and provide content to general online news sites such as Yahoo!. Universities and research institutions are also publicizing content directly to audiences through their own websites, social-media platforms and video channels.

## THE NEW NEWS

Young scientists tempted by a career in science journalism still need broad curiosity and a knack for writing, but there is also a growing need for digital media skills — including writing for and posting on the web, and the basics of web design — that can be learned on the job or through journalism training. Opportunities are not restricted to North America and Europe: writers who can critically understand and translate science-related stories about topics such as pollution or climate change are in high demand among the newly independent media outlets of the Middle East and Africa. Uncertainties remain about how science writing and media will evolve, but prospects are ►



► good for candidates who remain flexible and willing to pursue unconventional opportunities at unconventional outlets.

Robert Irion took over as director of the Science Communication Program at the University of California, Santa Cruz (UCSC), in 2007. Since then, he has seen 15 out of 52 graduates get jobs dedicated to online writing or producing. Science websites such as OurAmazingPlanet.com and Space.com are hiring full news teams; writers with biomedical training are finding jobs with websites such as the Alzheimer Research Forum (alzforum.org). Although positions are generally 'entry-level', says Irion, they offer dynamic environments and opportunities to move up the editorial ladder.

### ONLINE OUTLETS

"There are staff opportunities that didn't exist before," says Dan Fagin, director of SHERP. The scientific journal *The Proceedings of the National Academies of Science* launched a journalistic section this month; it has no staff positions apart from its editor, but has contracted 12 freelance writers. Quartz (qz.com), a global business-news website started last year and designed for delivery to tablets and smart phones, has hired writers for its energy and technology sections. Unlike their counterparts in earlier Internet booms, such web-only start-ups now appear to have more solid business plans and backing. Quartz is owned by Atlantic Media in Washington DC; the Alzheimer Research Forum is funded in large part by Banner Health, a non-profit health-care provider based in Phoenix, Arizona. It is not yet clear how secure the future of these sites will be in the long run, "but the market for our graduates is definitely better than it was even two to three years ago", says Fagin.

The picture is not bright everywhere. A popular dual master's programme in Earth and environmental science journalism at Columbia University in New York shut its doors to new applicants in 2010, mostly because of concern that graduates were struggling to find enough

work to pay off the substantial US\$89,000 tuition fee for the two-year degree, according to the *Columbia Journalism Review* (see [go.nature.com/htnw8j](http://go.nature.com/htnw8j)). A note on the university's website states that the programme will not be accepting applicants for the foreseeable future "due to the current weakness in the job market for environmental journalists".

Fagin says that in the past few years, an increasing number of SHERP graduates who want to be storytellers have gone straight into



**"Is the news conducive to a video? A feature? Or a tweet? If you can do all of that, you're going to be more employable."**

Melissa Lutz Blouin

are highly skilled and well connected.

Science-writing programmes now train students in a range of digital platforms and technologies. The courses at UCSC and New York University, for example, include work with digital photography, video and audio webcasting, blogging, social media and data-driven journalism — mining large, public data sets for science stories. Stephen Webster, director of the Science Communication Unit at Imperial College London, says that his programme, which offers two master's degrees, has blended

digital technologies into its five tracks of video documentary, radio broadcasting, writing, web design and gallery or museum exhibition.

"We caution incoming students, 'Don't imagine [the degree] means you'll be at a newspaper,'" says Webster. "If you are a good writer, it's more likely you'll be working on scripts for a radio documentary or in science communications." A 2012 poll that randomly sampled 251 alumni of the two programmes found that 44% were working in broadcast, print or online journalism and another 44% in science communications or related areas, including museums, policy and education.

The UCSC programme added social-media training because so many online positions require that writers and reporters be active across platforms such as Facebook, Twitter and Tumblr, says Irion. Hiring editors want writers who have built an online presence with thousands of followers, who know how to spread stories and "have a nose for news that has a potential to go viral", he says.

Fagin points out, however, that one thing has not changed much: a PhD is still by no means a prerequisite for a science-writing career. "Does it help? Sure. Is it absolutely necessary? Definitely not," he says. "Having the writing skills, journalistic instincts, facility with the tools and a deep understanding and appreciation of the scientific process are much more important."

### LINES OF COMMUNICATION

In the past decade, science-communication jobs have grown to outnumber conventional science-journalism posts in both North America and Europe. Hard numbers are difficult to come by, but in 2012, the US National Association of Science Writers in Berkeley, California, conducted a survey of its members. Out of 1,982 respondents, 16% reported holding staff positions at academic institutions, 3% at newspapers and 9% at magazines.

Christian Heuss, former chief of science staff at Swiss Radio and Television in Basel, says that in Switzerland, there are nearly twice as many

## GLOBAL OUTLOOK

### *An international snapshot of science writers*

Science writers' perceptions of the prospects for their field vary greatly by region, according to an international study by SciDev.net, a non-profit science-news organization based in London (M. W. Bauer *et al.* *Global Science Journalism Report* SciDev.net; 2013).

Using data from four different surveys — one conducted at the World Conference of Science Journalists in London in 2009, another by a Latin American team in 2010–11, and two by SciDev.net teams in 2012 — the organization built a database of responses from 953 science journalists.

Across the entire sample, about 72% of respondents are satisfied in their jobs. But only 29% in Europe and 32% in the United States and Canada would definitely encourage a young person to pursue science journalism. Other regions seem more positive, with 80% in North Africa and 72% in southern Africa reporting that they would encourage aspirants; the figures are 60% and 55% in Asia and Latin America respectively. Survey responses suggested that the pessimism in the United States and Canada is at least partly attributable

to concerns about the decline in print journalism.

Overall, 10% of science journalists who responded to the surveys hold a PhD, but that jumps to 32% in Europe and 31% in the United States and Canada. Full-time staff writers account for 51% of respondents, full-time freelancers 18% and part-time freelancers 14%. At least half of respondents use Twitter, other social media or blogs in their work, with numbers in the Middle East and North Africa higher than the global average. **K.P.**

major-university communication positions as there are conventional science-journalism jobs. Heuss himself has just made the switch to become head of communications at the Swiss Tropical and Public Health Institute in Basel.

The change is rooted in the growing demand for online content and a need for institutions to keep the public informed about taxpayer-funded research. Some publicly funded US universities are under hiring freezes, but many other institutions around the world are beefing up their communications teams.

Science writers working in academia need many of the same skills as science journalists, including a solid writing ability, research experience and digital know-how. Communications jobs have shifted from putting out the university's news of the day to targeting select audiences directly with marketing messages. Rather than solely writing press releases, science writers at universities now also craft speeches to donors and multimedia web presentations meant to highlight undergraduate research. Audiences include patients, students, faculty members, alumni, donors, legislators and decision-makers, as well as the regional, state and global public.

Borchelt says that research institutions are experimenting with how best to woo online audiences. "We can't take our traditional news products, like press releases, and just transfer them over to a new medium. [New media] work a lot differently from how old media worked," he says.

Writers must not only be comfortable with digital platforms and social media, but also be able to manage them wisely, says Melissa Lutz Blouin, director of news and publications for the University of Florida Academic Health Center in Gainesville. "Is [the news] conducive to a video or a slide show? A feature in a magazine that gets sent to decision-makers? Or is it short enough to do a tweet?" she asks. "If you can do all of that at some level, you're going to be more employable."

Recruiters stress the importance of being a collaborative 'people person', because a science writer will probably be part of a marketing team. "It's not yet a saturated market for people who can do both science and writing and do both well," says Borchelt.

## GOING GLOBAL

Conventional science journalism is not in decline everywhere: in the developing world, it is thriving (see 'An international snapshot of science writers'). Newly independent media in countries such as India, Venezuela and some African nations have growing, highly educated audiences that demand science coverage. Editors need journalists who can cover stories about the effects of environmental crises and how technology booms are aiding regional economies. Jean-Marc

Fleury, executive director of the World Federation of Science Journalists, who is based in Gatineau, Canada, notes that in the past few years, newspapers in Cameroon, Nigeria and Uganda have brought in science desks or pages. Internet access is still patchy in many of these areas, so print publications retain their appeal.

"In the past decade, Egypt has our first independent newspaper and now the region has a lot of publications that have a section



**"There are staff opportunities that didn't exist before."**

Dan Fagin

or a weekly column talking about science," says Bothina Osama, the Cairo-based Middle East and North Africa news editor for SciDev.net, a non-profit organization based in London that provides science news focused on the developing world. "That's given a nice boost in science-journalism jobs."

Osama and about 20 colleagues founded the Arab Science Journalists Association in 2002. The organization now has more than 230 members. "Since the Arab Spring, people are becoming more interested in reading about science because they see it as a force of socioeconomic advancement in the region. There is a political will to have more science news," says Osama.

In 2006, the World Federation of Science Journalists launched its Science Journalism Cooperation Project (SjCOOP), which seeks to raise the profile and standards of science journalism in Africa and the Arab world, and is funded by government international-development agencies from the United Kingdom, Canada and the Netherlands. The programme matches general journalists from this region with science-journalism mentors from Europe and North America. Editors-in-chief in the Middle East and Africa are asking for good journalists who can cover science, says Fleury.

But as elsewhere, Fleury says, would-be science journalists in the developing world should be open to jobs not clearly labelled 'science correspondent'. Irion suggests that young science writers must be go-getters in the job market. "I'm optimistic for any student who is willing to be entrepreneurial and wants to explore writing about science across different platforms," he says. "Yes, parts of the science-writing ecosystem have withered on the vine. But opportunities keep popping up." ■

**Kendall Powell** is a science writer based in Lafayette, Colorado.

## UNITED STATES

### Geospatial shortage

US agencies face a shortfall in employees who can visually depict and assess global security threats, finds a report released on 25 January by the National Academies in Washington DC. *The Future U.S. Workforce for Geospatial Intelligence* cites a need to measure Earth for surveying and environmental monitoring; determine objects' geometry through photos; and use visual data to synthesize information. It predicts personnel shortages at the US National Geospatial-Intelligence Agency owing to retirements, changing defence priorities and the need for international humanitarian support. Keith Clarke, a geographer at the University of California, Santa Barbara, and chairman of the report committee, suggests that interested scientists develop their skills in spatial thinking, mathematics and statistics.

## EUROPEAN UNION

### Inclusive rankings

The European Union on 30 January launched U-Multirank, a university ranking system based on broad criteria. In addition to research, used in most existing systems, it assesses universities' quality of teaching and learning; international networking; external partnerships; and regional engagement. Some 500 universities are expected to participate. A consortium led by the Centre for Higher Education in Gütersloh, Germany, and the Centre for Higher Education Policy Studies in Enschede, the Netherlands, will compile the first round of rankings by February 2014. The system is meant to improve transparency and reduce elitism in rankings, says a European Council representative.

## ENDOWMENTS

### University assets fall

Values of and returns from US university endowments have fallen, says a survey, but faculty recruitment is unlikely to be affected. The National Association of College and University Business Officers (NACUBO) in Washington DC and the Commonfund, an investment-management firm in Wilton, Connecticut, studied 831 institutions and found that endowments fell by an average of 0.3% between 2011 and 2012. Ken Redd, NACUBO director of research and policy analysis, says that faculty hiring is usually supported by other accounts and probably will not tighten as a result of the drop.

# PEACE

*A moment of doubt.*

BY DANNY DUNLAVEY

It's been 30 days since I killed my husband, my lover, my mate. I feel as though I should hide the evidence (although I don't think there is any), but it's all legal. Or rather, it's not illegal.

All the women know it's coming. The men don't. They find out during. Sort of.

When I asked Gary about his family, he told me that he only had a mother, as if that was natural. And about 100 brothers and sisters. He hardly knew about his father, meaning he didn't at all. Come to think of it, I don't know mine either.

Mum said not to worry about it. Murder comes naturally to the females of our species. It makes sense. We need the nutrients. I, of course, had no idea what she was talking about. She said I'd know when it was time.

And I did.

Martha came over yesterday. She tried to help console me.

"Julie, don't worry about it. I'm sure he didn't mind. Trust me, he died very happy," she said.

"Martha, I killed my husband! I can't just forget about that. I loved him, even if his forelimbs were a bit small. Size isn't what matters."

"Loved him? Love? Oh, never mind. Look, you'll get used to it. Do you know how many husbands I've had?"

"Twelve."

"And how many do I have now?"

"None."

"And am I all upset about it?"

"No, but you were always much more of an extrovert than me. When guys snapped their mandibles at me, my skin turned a red shade of green!"

"So? Even if your personality doesn't lend itself to mating, instinct will triumph. And then you'll learn that nothing's wrong," Martha said.

"Martha, it's not really even murdering Gary that I can't get over."

"Oh. Well, what are you all upset about? That's usually the hardest part. It gets so much more pleasurable after that. Put a few

spiders on it. It's delicious. Not the *sentient* spiders, of course."

"Martha, it's the cannibalism that

seems wrong to me."

"Ugh, I was worried you'd say that. Do you want to see Dr Pratt? A most excellent psychiatrist. Very considerate. When the kids get in trouble at school and I feel a stroke coming from the screaming, I pop a Valium and go comatose in front of the TV. The pinnacle of relaxation, trust me."

"OK. I'll go."

The office was nice. Dr Pratt was male. It looked like he was celibate, seeing as he was alive. He must have asked some serious questions about why he didn't have a father.

"Now, I must warn you, Miss Julie. It may be mating season, but I am just not interested," he began. "No offence. Your forelimbs and thorax are very nice."

"Thank you," I said. "You're perfectly attractive as well, but I think we can reach a mature agreement as two consenting and equal individuals that we should not have sexual relations, for both our sakes."

"Perfect," said Dr Pratt. "Now, Martha told me that you're having some guilt issues, but she didn't explain why. So ... what seems to be the problem?"

Great. Martha wanted to build up the maximum amount of embarrassment for me in front of a total stranger.

"My husband is dead, and I killed him. I know it's not illegal, but I feel there was some betrayal of trust when I cut his head off and ate him. I know it's irrational. I know, I know. But what can I do? I still love Gary," I said.

"I'm sorry, what? Did you say 'love'?" asked Dr Pratt.



"Yes. Love."

"... Right," he said hesitantly. "And what is this 'love' of which you speak?" he asked, moving his forelimbs up and down in the air.

"Well," I began, "I mean I thought Gary was handsome and sweet and he'd be a suitable mate."

"Oh. That's all. We usually just call that *desire* in the medical field. I suppose you may have seen that word somewhere in the old libraries or online," said Dr Pratt.

"You know what, you're right," I replied. "I've been reading books that some of the archaeologists have been putting online. 'Love' pops up a lot. I wasn't even thinking about it."

"Yes, yes. The civilization before the chemical spill a few centuries ago was dramatically different from ours. Suffice it to say, it was not very efficient."

"So, anyway," he continued, "why are you feeling guilty? It's not as if you weren't able to find a mate. You haven't let our species down."

"Well, Gary was more than just a suitable mate. He was sweet. He wrote me poetry, just like the stuff I've read online. He said we would be together forever."

"That does seem unusual for a mantis," said Dr Pratt.

"So, what do you think I should do? I can't ever love again, not after Gary. I mean, obviously I'll still raise all my children responsibly for the mantis race ..."

"Well, maybe you can find solace beyond the physical realm?"

"Beyond the physical?" I asked. "Like those irrational humans I've read about? Before the spill wiped them out? I mean, the poetry sounds nice, but the ideas behind it ..."

"Exactly. Just give it a try. I've read lots about the benefits in radical medical journals. You have a whole bunch of options: asking for forgiveness for stuff; 'spirituality'; coming to peace with yourself through inner dialogue ..."

"All right," I said, "if you think it'll help."

"Now, let's see here ..." he began flipping through a book on his desk. "Right. Just like the name given to our ancestors in the old texts — I believe it's called 'praying.'" ■

➔ **NATURE.COM**  
Follow Futures:  
@NatureFutures  
go.nature.com/mtoodm

**Danny Dunlavy** is a writer from Maryland. He talks fast, punches trees and murders rocks. In addition, he watches movies and reads. Yes, he knows he's not interesting.

JACEY